A Longitudinal Study of Academic Web Links: Identifying and Explaining Change

Nigel Payne BSc(Hons), MSc

A thesis submitted in partial fulfilment of the requirements of the University of Wolverhampton for the degree of Doctor of Philosophy

January 2008

This work or any part thereof has not previously been presented in any form to the University or to any other body whether for the purposes of assessment, publication or for any other purpose (unless previously indicated). Save for any express of acknowledgements, references and/or bibliographies cited in the work, I confirm that the intellectual content of the work is the result of my own efforts and of no other person.

Signature

Date

Acknowledgements

First of all, I would like to thank my supervisor, Professor Mike Thelwall. He has been my guide and mentor from the early days of my Masters project, consistently providing invaluable support and advice for the past four years. I would like to thank Helen Ashdown for initially encouraging me to undertake a PhD and arranging my first meeting with Mike. I think myself fortunate to have found Mike at a time where my studies had no clear direction. During my research, he has struck a perfect balance between providing subtle guidance, both technically and professionally, and giving me the space to generate my own ideas. His tireless encouragement and assistance at every stage of my studies has been inspiring and has allowed me to gain confidence in my subject. It has been a real privilege for me to have him as a supervisor.

Furthermore, I wish to express my gratitude to the members of the University of Wolverhampton Statistical Cybermetrics Research Group who have provided constant support and constructive feedback. As I have conducted most of my studies at some distance from the University, regular contact with these individuals has undoubtedly helped me with my research and made my visits both productive and enjoyable.

Last, but certainly not least, I would like to thank my wife, Christina. She has been incredibly supportive and patient throughout my studies and I am absolutely certain that without her influence and confidence, I would not have reached this stage of my PhD.

I dedicate this thesis to my grandparents, Jim and Evelyn Crimmings. They were, and still are, a major positive influence in my life, and are deeply missed.

Publication List

PAYNE, N. & THELWALL, M. (2004) A Statistical Analysis of UK Academic Web Links. *Cybermetrics*, **8**(1), paper 2. Available: http://www.cindoc.csic.es/cybermetrics/articles /v8i1p2.html

PAYNE, N. & THELWALL, M. (2005) Mathematical Models for Academic Webs: Linear Relationship or Non-Linear Power Law? *Information Processing and Management*, **41**(6), pp. 1495-1510.

THELWALL, M. & PAYNE, N. (2005) Link Analysis: An Informetric Technique. *Proceedings of the 10th International Conference on Scientometrics and Informetrics*, *ISSI-2005*. pp. 681-682.

PAYNE, N. & THELWALL, M. (2007a) A Longitudinal Study Of Academic Webs: Growth and Stabilisation. *Scientometrics*, **71**(3), pp. 523-539.

PAYNE, N. & THELWALL, M. (2007b, to appear) Longitudinal Trends in Academic Web Links. *Journal of Information Science*.

PAYNE, N. & THELWALL, M. (2007c, to appear) A Longitudinal Analysis of Alternative Document Models. *ASLIB Proceedings.*

PAYNE, N. & THELWALL, M. (2007) Do Academic Link Types Change Over Time? has been submitted to the *Journal of Documentation*.

Although the publications listed above were published in collaboration with Professor Mike Thelwall, the candidate's individual contribution to the content of the first-named papers was 100%, with Professor Thelwall acting in a supervisory and advisory capacity (and as a second classifier in the 'Do Academic Link Types Change Over Time?' paper).

As well as the list of publications above, the candidate has also served as a referee for the Journal of Information Science.

Abstract

A problem common to all current web link analyses is that, as the web is continuously evolving, any web-based study may be out of date by the time it is published in academic literature. It is therefore important to know how web link analyses results vary over time, with a low rate of variation lengthening the amount of time corresponding to a tolerable loss in quality. Moreover, given the lack of research on how academic web spaces change over time, from an information science perspective it would interesting to see what patterns and trends could be identified by longitudinal research and the study of university web links seems to provide a convenient means by which to do so.

The aim of this research is to identify and track changes in three academic webs (UK, Australia and New Zealand) over time, tracking various aspects of academic webs including site size and overall linking characteristics, and to provide theoretical explanations of the changes found. This should therefore provide some insight into the stability of previous and future webometric analyses.

Alternative Document Models (ADMs), created with the purpose of reducing the extent to which anomalies occur in counts of web links at the page level, have been used extensively within webometrics as an alternative to using the web page as the basic unit of analysis. This research carries out a longitudinal study of ADMs in an attempt to ascertain which model gives the most consistent results when applied to the UK, Australia and New Zealand academic web spaces over the last six years. The results show that the domain ADM gives the most consistent results with the directory ADM also giving more reliable results than are evident when using the standard page model. Aggregating at the site (or university) level appears to provide less consistent results than using the page as the standard unit of measure, and this finding holds true over all three academic webs and for each time period examined over the last six years.

The question of whether university web sites publish the same kind of information and use the same kind of hyperlinks year on year is important from the perspective of interpreting the results of academic link analyses, because changes in link types over time would also force interpretations of link analyses to change over time. This research uses a link classification exercise to identify temporal changes in the distribution of different types of academic web links, using three academic web spaces in the years 2000 and 2006. Significant increases in ‘research oriented’, ‘social/leisure’ and ‘superficial’ links were identified as well as notable decreases in the ‘technical’ and ‘personal’ links. Some of these changes identified may be explained by general changes in the management of university web sites and some by more wide-spread Internet trends, e.g., dynamic pages, blogs and social networking. The increase in the proportion of research-oriented links is particularly hopeful for future link analysis research.

Identifying quantitative trends in the UK, Australian and New Zealand academic webs from 2000 to 2005 revealed that the number of static pages and links in each of the three academic webs appears to have stabilised as far back as 2001. This stabilisation may be partly due to an increase in dynamic pages which are

normally excluded from webometric analyses. In response to the problem for webometricians due to the constantly changing nature of the Internet, the results presented here are encouraging evidence that webometrics for academic spaces may have a longer-term validity than would have been previously assumed.

The relationship between university inlinks and research activity indicators over time was examined, as well as the reasons for individual universities experiencing significant increases and decreases in inlinks over the last six years. The findings indicate that between 66% and 70% of outlinks remain the same year on year for all three academic web spaces, although this stability conceals large individual differences. Moreover, there is evidence of a level of stability over time for university site inlinks when measured against research. Surprisingly however, inlink counts can vary significantly from year to year for individual universities, for reasons unrelated to research, underlining that webometric results should be interpreted cautiously at the level of individual universities.

Therefore, on average since 2001 the university web sites of the UK, Australia and New Zealand have been relatively stable in terms of size and linking patterns, although this hides a constant renewing of old pages and areas of the sites. In addition, the proportion of research-related links seems to be slightly increasing. Whilst the former suggests that webometric results are likely to have a surprisingly long shelf-life, perhaps closer to five years than one year, the latter suggests that webometrics is going to be increasingly useful as a tool to track research online.

While there have already been many studies involving academic webs spaces, and much work has been carried out on the web from a longitudinal perspective, this thesis concentrates on filling a critical gap in current webometric research by combining the two and undertaking a longitudinal study of academic webs. In comparison with previous web-related longitudinal studies this thesis makes a number of novel contributions. Some of these stem from extending established webometric results, either by introducing a longitudinal aspect (looking at how various academic web metrics such as research activity indicators, site size or inlinks change over time) or by their application to other countries. Other contributions are made by combining traditional webometric methods (e.g. combining topical link classification exercises with longitudinal study) or by identifying and examining new areas for research (for example, dynamic pages and non-HTML documents).

No previous web-based longitudinal studies have focused on academic links and so the main findings that (for UK, Australian and New Zealand academic webs between 2000 and 2006) certain academic link types exhibit changing patterns over time, approximately two-thirds of outlinks remain the same year on year and the number of static pages and links appears to have stabilised are both significant and novel.

Table of Contents

List of Figures

List of Tables

1 General Introduction

1.1 Introduction

Link analysis is concerned with identifying relationships and associations between objects not apparent from isolated pieces of information. Link analysis is carried out in many different disciplines, including much recent social science research conducted with the aim of analysing the social interactions between individuals. In the context of this study, we are concerned with *web* link analysis, which is carried out in order to understand and extract information from the link structure of collections of web documents. A distinct feature of the web is the use of hyperlinks between web pages which allow a user to navigate from one web page to another. The underlying assumption is that these links contain useful information and web link analysis seeks to uncover the relationships and patterns within this hyperlink data.

Web link analysis has already established its own methodologies and tools and these have been used for research in many other disciplines including library science, computer science, theoretical physics and sociology. This is indicative of a widespread belief, driven by the obvious importance of hyperlinks on the web, that links between web pages can yield useful information. The current study hopes to provide evidence to support this belief, by showing how academic links have changed and developed over time.

Web link analysis is a subset of the field of webometrics (itself a subsidiary of bibliometrics) and has mainly focussed on the transference of traditional informetric techniques to web-based data. Many previous studies have provided significant results using analogies such as journal citations to hyperlinks and Journal Impact Factors to Web Impact Factors. A large body of research has been dedicated to the examination of scholarly communication on the web and, although the dissemination of scholarly information was initially a main driving force for the creation of the Internet, it has developed in size, scope and importance into a primary global information repository, communication and research tool.

1.2 Research Motivation

As the review of literature in chapter 2 shows, there have already been many webometric studies of academic webs. It can also be seen that the Internet has provided a natural area of interest for much longitudinal research. However, there has been no specific longitudinal study of academic web spaces, which is the critical gap the current study hopes to fill. The need for conducting a longitudinal study of this nature has stemmed from the realisation that the quality of web-based research may diminish over time (as highlighted by Björneborn, 2004), and that knowledge of the stability of web documents and links would be critical to ensure any relevance to webometric studies.

A problem common to all current web link analyses is that, as the web is continuously evolving, any web-based study may be out of date by the time it is published in academic literature. Therefore, it is important to know how web link analyses results vary over time with a low rate of variation lengthening the amount of time corresponding to a tolerable loss in quality. Moreover, given this lack of research on how academic web spaces change over time, from an information science perspective, it would interesting to see what patterns and trends could be identified by longitudinal research and the study of university web links provides a convenient means by which to do so.

Since this study centres around, and makes frequent use of the term ‘web space’ it is important to first define this term. In the absence of any formal webometric description, this study proposes the following definition: ‘a collection of related web pages, (sub)domains or web sites’. This is an abstract idea, not necessarily tied to any specific location but rather ‘somewhere in cyberspace’. An academic web space would necessarily include educational institutions such as universities or research centres. A very large academic web space may be spread out over a number of web servers in different geographical locations. In the context of this study, the academic web spaces for the UK, Australia and New Zealand incorporate all identified university web pages, directories, subdomains, domains and sites for each country.

Academic web spaces, as an area of study for this investigation, were chosen for the following reasons:

- The Internet was initially created for scholarly use and has since developed into an essential conduit for both formal and informal scholarly communication and collaboration. Academic websites may provide an opportunity for close comparison with academic research.
- Academic web spaces generally appear to employ consistent domain naming protocols, which greatly aid accurate data collection and processing.
- There exists a current, comprehensive data set provided by the University of Wolverhampton Academic Web Link Database Project which contains text files of the hyperlink structure for each UK, Australian and New Zealand university for the last 6 years.
- The number of universities, (125 UK, 38 Australia and 8 New Zealand), is manageable, yet large enough for meaningful statistical measure to be taken.
- The academic web is mature, and appears to be relatively well interlinked.
- Academic webs have already been the subject of extensive recent research and so a study of how and why they change would be both informative and relevant to current studies.
- There should be little financial or business oriented motivation for linking between universities (Bar-Ilan, 2004b). (This may not be entirely true as academic institutions may choose to ignore the competition. However, if it is true, this setting enables us to study ‘pure’ linking motivations).
- Henzinger, Motwani and Silverstein (2002) discuss ‘web conventions’ (rules which web authors follow without anybody formally imposing these rules on them). If such conventions do exist, then they should appear in the academic environment.

1.3 Research Aims and Objectives

1.3.1 Research Aims

The aim of this research is to identify and track changes in three academic webs (UK, Australia and New Zealand) over time, using quantitative techniques, tracking various aspects of academic webs including site size and overall linking characteristics, and to provide theoretical explanations of the changes found. The purpose of this research is to provide some insight into the stability of results for academic webometric studies.

1.3.2 Research Objectives

The objective of this research is to answer the following primary Research Question:

How have the UK, Australian and New Zealand academic webs changed over time since 2000 and what are the implications for the stability of webometrics research?

In order to answer this question, this research will concentrate on the following more specific sub-questions:

Which Alternative Document Models give the most consistent results when applied to the UK, Australian and New Zealand academic web spaces 2000 – 2006?

What types of UK, Australian and New Zealand academic web links change over time and why?

What is the trend over time for the average web site size, and average inlink count, of UK, Australian and New Zealand academic webs?

How has the relationship between university inlinks and research activity indicators varied over time?

Which universities in each of the UK, Australian and New Zealand academic webs have experienced the greatest change in inlinks and outlinks over the last six years, and why?

1.4 Structure of Thesis

Chapter 2 carries out a comprehensive review of relevant literature in order to put the current research into context, and to identify any critical gaps. The review begins by positioning webometric research within the wider framework of informetric, bibliometric and scientometric study. Webometrics is broken down into web page content analysis, web technology analysis, web usage analysis and web link structure analysis and a critical review of previous research in each field is undertaken.

The current study is positioned firmly within the scope of web link structure analysis and specifically, academic web link structure analysis. Previous research is critiqued using national, international and departmental university link analysis as areas of interest, and established webometric tools such as the Web Impact Factor and Alternative Document Models are discussed.

Although not directly related to the current study, web topology and social network analysis are examined to provide essential background information. Specific Internet longitudinal studies are then discussed to provide a historical perspective to the current research. Limitations, both generic to web-based research and specific to this study, are identified and reviewed throughout.

Chapter 3 deals with the research design, outlining the research approach and research philosophy before stating, and justifying, the research hypothesis. Chapter 4 highlights the research methods used, detailing data collection methods such as the SocSciBot Crawler and Tools, the University of Wolverhampton Academic Web Link Database Project, bespoke tools and the Internet Archive. The use of research and academic staff data are also discussed.

The main body of the thesis consists of specific empirical investigations into identifying changes and trends in university web link data and is broken down into the chapters described below.

Previous studies of academic web interlinking have tended to hypothesise that the relationship between the research of a university and links to or from its web site should follow a linear trend, and yet the typical distribution of web data, in general, seems to be a non-linear power law. Hence it was necessary to formally establish whether a linear trend or a power law was the most appropriate method with which to model the relationship between research and web site size or outlinks, and this is dealt with in chapter 5. Following linear regression, analysis of the confidence intervals for the logarithmic graphs, and analysis of the outliers, the results suggest that a linear trend is more appropriate than a non-linear power law and, in the remainder of the thesis it is assumed that a linear model is best for the data. The main content of this chapter has previously been published as 'PAYNE, N. & THELWALL, M. (2005) Mathematical Models for Academic Webs: Linear Relationship or Non-Linear Power Law? *Information Processing and Management*, **41**(6), pp. 1495-1510'.

Alternative Document Models (ADMs) have been used extensively within webometrics as an alternative to using the web page as the basic unit of analysis. They were created with the purpose of reducing the extent to which anomalies occur in counts of web links at the page level. Chapter 6 has been accepted for publication as 'PAYNE, N. & THELWALL, M. (2007c) A Longitudinal Analysis of Alternative Document Models. *ASLIB Proceedings*' and carries out a longitudinal study of ADMs in an attempt to ascertain which model gives the most consistent results when applied to the UK, Australia and New Zealand academic web spaces over the last six years. The results show that the domain ADM gives the most consistent results. However, the directory ADM also gives more reliable results than are evident when using the standard page model. Aggregating at the site (or university) level appears to provide less reliable results than using the page as the standard unit of measure, and this finding holds true over all three academic webs and for each time period examined over the last six years.

The main content of chapter 7 has been submitted for publication in the *Journal of Documentation* and addresses the question of whether university web sites publish the same kind of information and use the same kind of hyperlinks year by year or whether these change over time. This is an important issue from the perspective of interpreting the results of academic link analyses, because changes in link types over time would also force interpretations of link analyses to change over time. This chapter uses a link classification exercise to identify temporal changes in the distribution of different types of academic web links, using the academic web spaces of the UK, Australia and New Zealand in the years 2000 and 2006. Significant increases in 'research oriented', 'social/leisure' and 'superficial' links were identified as well as notable decreases in the 'technical' and 'personal' links. Some of these changes identified may be explained by general changes in the management of university web sites and some by more wide-spread Internet trends, e.g., dynamic pages, blogs and social networking. However, the increase in the proportion of research-oriented links is particularly hopeful for future link analysis research.

Most of the content of chapter 8 has been previously published as 'PAYNE, N. & THELWALL, M. (2007a) A Longitudinal Study Of Academic Webs: Growth and Stabilisation. *Scientometrics*, **71**(3), pp. 523-539' and studies the UK, Australian and New Zealand academic webs from 2000 to 2005, finding that the number of static pages and links in each of the three academic webs appears to have stabilised as far back as 2001. This stabilisation may be partly due to an increase in dynamic pages which are normally excluded from webometric analyses. In response to the problem for webometricians due to the constantly changing nature of the Internet, the results presented in this chapter are encouraging evidence that webometrics for academic spaces may have a longer-term validity than would have been previously assumed.

Chapter 9 examines the relationship between university inlinks and research activity indicators over time and identifies reasons for individual universities experiencing significant increases and decreases in inlinks over the last six years. The findings also indicate that between 66% and 70% of outlinks remain the same year on year for all three academic web spaces, although this stability conceals large individual differences. Moreover, there is evidence of a level of stability over time for university site inlinks when measured against research. Surprisingly however, inlink counts can vary significantly from year to year for individual universities, for reasons unrelated to research, and this undermines their use in webometric studies.
The main findings of this chapter have been accepted for publication as 'PAYNE, N. & THELWALL, M. (2007b, to appear) Longitudinal Trends in Academic Web Links. *Journal of Information Science*'.

Chapter 10, the discussion, outlines the significant contributions which the current study has made to the existing knowledge base as well as highlighting some of the recurring trends identified within this research. It also attempts to isolate some examples of where the thesis fulfils certain criteria for originality.

Chapter 11 presents a conclusion to the present work, as well as outlining some possibilities for future research.

1.5 Summary

As well as outlining the structure of the thesis above, this chapter also attempts to clarify the motivation for the research, in addition to listing its aims and objectives. The themes that emerge from this research are valuable in a number of respects; longitudinal analyses of the link structure of the academic web should highlight patterns and trends which show how the academic web has developed and should provide some indication of the duration of the validity of web-based research.

This research is important because a greater understanding of the longitudinal relationships and patterns within the hyperlink structure of the academic web will develop an appreciation of the way this web is currently connected and may prove to be useful in predicting future development and evolution.

2 Literature Review

2.1 Introduction

The purpose of this chapter is to examine relevant literature in the field with a view to summarising the important papers and giving an indication of how the current study has developed from previous work.

This literature review is a systematic review and critique of the published work relevant to the current study and has the following main functions:

- It places the current study in a historical perspective
- It shows how the research fits with other work in the subject area
- It acts as a guide and reference to further reading in the area
- It indicates alternative views

It is important to commence all research with a review of the related literature in order to determine whether any data sources already exist that can be brought to bear on the problem at hand; this is also referred to as secondary research. The quality of the literature being reviewed must be carefully assessed as not all published material is the result of good research design or can be substantiated. This literature review attempts to be critical of what has been written, identify areas of controversy and identify areas which need further research. In addition to undertaking an analysis of relevant literature, this chapter includes critical analyses of research methods employed where appropriate e.g. web crawlers, search engines and classification methodologies.

In an attempt to establish the context of this research, this chapter outlines the history of the development of webometrics and positions the research within the wider field incorporating informetrics, bibliometrics, scientometrics and cybermetrics.

2.2 Informetrics and Early Quantitative Web Studies

Tague-Sutcliffe (1992) defined informetrics as the study of the quantitative aspects of information in any form, not just records and bibliographies, and in any social group, not just scientists. It can therefore incorporate, utilise and extend studies of the measurement of information that lie outside the boundaries of both bibliometrics (the study of the quantitative aspects of documents) and scientometrics (the study of the quantitative aspects of science and its outputs) (Leydesdorff, 2001). Earlier, using aspects from both bibliometrics and scientometrics, Garfield (1955) had introduced citation indexes for scientific literature which enabled detailed analyses of citation networks. This was later followed by the introduction of the co-citation, the frequency with which two documents are cited together (Small, 1973). Citation analysis is a well established technique that has been successfully carried out on a variety of subjects including journals (Frandsen, 2005), academic papers (Redner, 1998), authors (Goodall, 2006) and article titles (Uzun, 2006).

The growing popularity of the Internet in the mid 1990s, and its use as a primary source of information provided an obvious research field for bibliometricians and scientometricians, and prompted a range of studies concerned with transferring established informetric methods to the web. Bossy (1995) appears to have been the first researcher to spot the potential to apply established information science techniques to the Internet. Then, in his exploratory analysis of the intellectual structure of cyberspace, Larson (1996) explicitly adapted existing information science techniques from bibliometrics to the web with the objective of assessing the link structure of a specific topic (Earth Sciences) and the characteristics of highly linked-to documents.

Following this, other information scientists realised that the advanced features of search engines could be used for information science link analyses. This produced Downie's (1996) application of informetric modelling techniques to the web, Turnbull's (1996) attempts to show the relevance of using operational bibliometric methods on the world wide web, Rousseau's (1997) informetric analysis of the web (using the term 'sitation' as the web analogy of 'citation') and Rodriguez i Gairín's (1997) web citation analysis, in which he describes the search engine AltaVista as the web's 'citation index'.

A natural extension of citation analysis was its application to e-journals. Initial results were disappointing with Smith (1999) using AltaVista to count links to 22 Australian refereed e-journals from the rest of the web. No significant relationships were found during this research with the author claiming that hyperlinks to e-journals are fundamentally different to citations as the former target the whole journal whereas the latter target individual articles. Harter and Ford (2000) reached similar conclusions for a set of 39 journals, also finding no significant correlation between link measures and Institute for Scientific Information (ISI) impact factors. Vaughan and Hysen (2002) went on to analyse an ISI-indexed set of library and information science journals. The journals in their study were not e-journals in the sense of providing full text of articles online but traditional journals with associated web sites. This time a significant correlation was found between inlink counts and Impact Factors which may be explained by the greater maturity of the web by that time and the selection of journals from a single discipline.

Lawrence, Giles and Bollacker (1999) took a different approach to citation analysis on the web. Instead of studying hypertext links as analogies of citations in the academic world, they looked for citations in the classical sense. Their 'Autonomous Citation Indexing' (ACI) system seems effective in automatically creating a citation index from electronic research material posted on the web, although its user interface has attracted some criticism.

The relevance of adopting and adapting existing information science techniques to the analysis of Internet documents can also be demonstrated by their successful application to electronic information retrieval systems and online databases (Wolfram, 1992; 2000; Cooley, Mobasher & Srivastava, 1997; Khan & Locatis, 1998; Wormell, 1998; Chakrabati et al., 1999; Bar-Ilan, 2001; Heinzinger, 2001; Nelson & Downie, 2002; Ajiferuke & Wolfram, 2004a; Nicholson, 2006; Beaulieu & Simakova, 2006).

However, while Cronin (2001) and Van Raan (2001) acknowledge the opportunity for bibliometricians to apply old techniques to new contexts, and Vaughan and Shaw (2003) establish some correlation between bibliographic and web citations, Egghe (2000) cautions on their direct comparison. Although web page outlinks and inlinks could be considered as similar to references and citations in scientific articles, the analogy drawn between classical citations and hypertext links is dangerous. For instance, a journal paper 'B' citing paper 'A' was necessarily written after paper 'A' but this is not necessarily the case with web pages (Rousseau & Thelwall, 2004 expand further on this). Quite often there are reciprocal links between web pages and these do not have analogies with citations. Also, as citations are taken from refereed journals they could be considered to have been subjected to some kind of quality control, whereas hyperlinks are not (Prime, Bassecoulard & Zitt, 2002).

Therefore, although the application of informetric techniques to the Internet offers the potential to uncover hidden patterns and relationships (Egghe, 2005; 2006), its results should be interpreted with caution and Bjorneborn and Ingwersen (2004) warn of the dangers in relying on the analogy between citation analyses and link analyses. Indeed, in an exploratory study, Smith (2004) found that only 20% of links to research oriented web sites could be regarded as research links analogous to citations.

2.3 Webometrics

Webometrics, the quantitative study of web-related phenomena, originated from the realisation that methods originally designed for bibliometric analysis of scientific journal article citation patterns could be applied to the web, with commercial search engines providing the raw data. Popular search engines, initially AltaVista, made available a mechanism which calculated link counts between web spaces using Boolean searches.

Almind and Ingwersen (1997) defined the field and coined the term 'webometrics' for the quantitative analysis of web-related phenomena from an information science perspective. A year later, Aguillo founded the e-journal using the name 'Cybermetrics'. These two terms are widely used as synonyms although cybermetrics includes quantitative analysis of the Internet, not just of the web (Aguillo, 2002). Bossy (1995) had previously introduced the term 'netometrics' but few information scientists have since taken up its use. Since its introduction, a number of academic papers have attempted to formally define the term 'webometrics', and to position it within the traditional information science framework (Björneborn & Ingwersen, 2001; 2004; Kretschmer & Thelwall, 2004; Thelwall, Vaughan & Björneborn, 2005).

Many different types of webometric research have been conducted and these include Cui (1999) and Hernandez-Borges, Macias-Cervi and Gaspar's (1999) use of webometric techniques to rate the relative quality of Internet health sites, while Thomas and Willet (2000) discovered that sitation data is not well suited to the quantitative evaluation of the research status of Library and Information Science (LIS) departments. Foot et al., (2006) analysed linking practices for US electoral candidate's web sites in the year 2002, while Kretschmer, Kretschmer and

Kretschmer (2007) examined web hyperlinks and web visibility indicators to establish their usefulness as indicators of collaboration and to explore whether similarities exist between web-based structures and bibliographic structures. Thelwall and Wilkinson (2003a) and Bar-Ilan (2004a) successfully compared established bibliometric techniques such as couplings, co-linking and self-linking to their analogies within academic web spaces. Park and Thelwall (2003) undertook a review and comparison of webometrics and hyperlink network analysis (derived from social network analysis) concluding that although both approaches examine the relational attributes between web sites, webometrics tends to concentrate on hyperlinks between academic web sites while hyperlink network analysis tends to view hyperlinks as social and communication ties.

Web based data, if sufficiently understood, have the potential to produce a new generation of science and technology (Aguillo, 1998) or financial (Uberti & Maggioni, 2004) indicators for the digital era, as well as supporting a range of social science research (e.g., Garrido & Halavais, 2003; Park, Barnett & Nam, 2002).

Thelwall, Vaughan and Björneborn (2005) proposed four main sub-areas for webometric research; web page content analysis, web technology analysis, web usage analysis and web link structure analysis and these are used during this literature review.

2.3.1 Web Page Content Analysis

Bar-Ilan and Peritz (2002) state that 'content analyses of web and Internet sources serve as exploratory tools for getting a better understanding of the Internet's content'. The realisation that web pages with common content often link to each other in an attempt to guide readers to related material on other web sites gives information scientists the opportunity to investigate patterns inherent in this structure of social navigation.

Early web content analyses include Rosenbaum's (1998) analysis of the content of the web sites of 24 web-based community networks in Indiana in an attempt to learn about the content and the structure of these sites. Haas and Grams (2000) identified seven different page types using a content analysis of 331 randomly selected web pages. However, while many of the pages could be clearly identified as one type, others were found to be combinations of several types. Bar-Ilan (2000) analysed the content of web pages containing the phrase 'S&T indicators' finding a large number of these web pages originated in Malaysia.

While Weare and Lin (2000) acknowledged the importance of web-based content analysis, they realised that it presents challenges as well as opportunities. They stated that content-based studies may be missing an opportunity if they failed to consider the hyperlinked environment, and went on to suggest methodological improvements in an attempt to improve future Internet content-based research.

Menczer (2001) attempted to validate the 'link-content conjecture' which states that the lexical content of a page can be inferred by looking at the pages that link to it. The results of this study could explain the success of the search technologies in use at that time. Chakrabarti et al., (2002) agreed and proposed that a topic taxonomy such as Yahoo! could be used as a framework for understanding the structure of content-based clusters and communities on the web. They introduced a measure of the probability that a page about one broad topic will link to another broad topic.

There is also a large body of work on the automatic classification of web pages based on their content using machine learning algorithms. Sebastiani (2002) discusses the main approaches to text categorisation which fall within the machine learning paradigm while web-based approaches include Yang, Slattery and Ghani (2002) who used three hypertext datasets and three well-known learning algorithms to examine regularities in different domains and compared different ways to exploit them. Dumais and Chen (2000) explored the use of hierarchical structure for classifying a large, heterogeneous collection of web content to support classification of search results, adding to a growing body of work exploring how hierarchical structures can be used to improve the efficiency and efficacy of text classification while Glover et al. (2002) analysed the relative utility of document text, and the text in citing documents near the citation, finding that the text in citing documents often has greater discriminative and descriptive power than the text in the target document itself.

2.3.1.1 Page Classification

Web researchers have long realised that classification of the various categories of web pages could lead to a better understanding of the relative importance of different types of pages in the structure of the web. During his Master's thesis, Almind (1995) introduced an early example of classifying web pages according to the function given to them by their authors as follows:

Personal Home Page: a home page whose main purpose is to represent an individual.
Institutional/Organisational Home Page: a home page whose main purpose is to represent an organisation.
Subject Defined/Ad Hoc Home Page: a home page whose main purpose is to represent a subject.
Pointer Document/Index Page: a web page whose function is primarily to make a number of hyperlinks available.
Resources: web pages which primarily make data available, for example, in the form of text, sound, pictures or film.

Cronin et al., (1998) searched the web for pages mentioning five prominent LIS professors using five popular search engines. The results were then characterised into one of the following eleven 'forms of mention'; abstract, article, conference proceedings, current awareness, external home page, listserv, personal/parent organization home page, resource guide, book review, syllabus or table of contents.

Several other studies have used web links, either solely or as part of a more integrated solution, to classify web pages. Furnkranz (1998) used web links to classify pages on the assumption that it was easier to classify hypertext pages using information on pages that point to a page instead of using information provided on the page itself. Calado et al., (2005) evaluated how the link structure of the web can be used to determine a measure of similarity appropriate for document classification. Tests performed on a web directory showed that link information alone allows for the classification of documents with an average precision of 86% and, when combined with a traditional text-based classifier, precision increased to values of up to 90%. Koehler (1999c) offers three separate approaches to aid the classification of web documents including the use of URL characteristics, while Harrison (2002) explored the semantic and rhetorical principles underlying link creation and proposed a classification of links according to their primary function. Crowston and Williams (2000) used a random sample of 1000 URLs from the AltaVista search engine to classify around 50 different genres on the web.

2.3.1.2 General and Commercial Link Creation Motivation Studies

Much web link-based research has aimed to identify patterns and relationships within hyperlink structures. However, considering the importance of identifying the reasons for constructing these links in the first place, studies into the motivation behind the creation of hyperlinks are relatively scarce. Some social network analyses tend to assume that the motivations of hyperlinking to another person's or institution's web site are recommendations or endorsements of the site but hyperlinks may have negative connotations also (Sunstein, 2001). Motivational studies are vital in order to develop an understanding of how link counts should be interpreted.

Kim (2000) investigated motivations for creating links in electronic publications in order to examine the analogy between citations and scholarly linking in electronic journals. Scholarly, social and technological reasons were identified, with most links having more than two reasons as motivation for creation.

Harrison (2002) explored the semantic and rhetorical principles underlying link creation and proposed a classification of links according to their primary function. Park (2002) conducted a survey of 64 Korean webmasters of commercial web sites to assess their motivation for linking to other web sites. He found that the webmasters were more likely to hyperlink to web sites possessing practical content, information or services. He also found that, although web links were generally created for either navigational functionality or business purposes, webmasters require that the credibility of hyperlinked web sites be higher than average when deciding to hyperlink to them.

Park, Barnett and Nam (2002) took an alternative view and regarded the number of inlinks to a web site as an indicator of site credibility. They empirically tested associations among hyperlink network structures, the number of visitors and Internet users perceptions of the web sites credibility, using a sample composed of South Korean web sites, and found that a site's incoming centrality in the hyperlink network was significantly related to visiting behaviour and perceived web site credibility.

2.3.1.3 Academic Page and Link Classifications

A number of web-based studies have undertaken academic link classification exercises. Chu (2005) analysed a random sample of links from academic institution's web sites and generated a list of reasons for hyperlinking. On the whole, almost 50%

of all inlinks examined were created for pointing to resource or directory information provided at the target web sites. In addition, 73% of all the inlinked sites analysed were linked to for reasons relating to service or home page while less than one third (27%) of the links were made out of research or teaching/learning motivations. Thelwall (2003c) took a sample of 100 random inter-site links to UK university home pages and postulated 4 new types of motivation; ownership, social, general navigational and gratuitous. A comparable study was undertaken by Kousha and Horri (2004) for Iranian universities finding that, of the 440 links studied, 63% were made for gratuitous or navigational reasons. Links between UK universities have also been analysed using a three-faceted (mathematics, physics and sociology) link source and target categorization scheme (Harries et al., 2004).

The most comprehensive academic hyperlink motivation study so far took a random collection of 414 links between UK university web sites and classified them according to the apparent motivation for their creation by two independent researchers (Wilkinson et al., 2003). The classification was problematic due to a low level of agreement on categories. Indeed, link creation motivation is a subjective issue and various judgements can be extracted from reasons for creating links (Kousha, 2005). Nevertheless, by combining similar categories, more reliable ones were formed and it was shown that although less than 1% of hyperlinks targeted formal scholarly publications such as a journal article or conference paper, over 90% of targeted material was in some way related to research or other scholarly activity such as teaching. Even when the motivations for link creation are known, concerns still remain about the validity of using academic web links as analogues to citations and this study shows that web hyperlinks are best viewed as data about informal scholarly communication. Bar-Ilan's (2004b; 2004c) academic link studies include categories for the type of source and target page of inter-university links in Israel. Although the categories and data collection methods used by these studies and Wilkinson et al., (2003) are different (a commercial engine and a specialist web crawler respectively), the results are similar. Bar-Ilan's (2004b) 'research-related' category contains 20% of all links studied while 27% of links in Wilkinson et al.'s (2003) classification fall into the 'research support and resources', 'research partners' or 'research reference' classes. Excluding 'superficial' and 'technical' links from Bar-Ilan's (2004c) study shows that approximately 86% of links are related to scholarly activity, reflecting the findings of Wilkinson et al., (2003).

A number of studies have attempted to classify university pages as opposed to links. Thelwall (2001d) carried out a classification of pages in an attempt to differentiate between research oriented and non-research oriented links. A later study of the top 100 most highly linked-to pages in UK universities found no recreational pages in the data set, instead finding it dominated by university home pages (Thelwall, 2002i). The most detailed subject classification applied to academic web sites is that of the UK's national RAE, consisting of 68 subject categories (Thelwall, Harries & Wilkinson, 2003) in which a random sample of 586 pairs of interlinked domain name-based web sites were classified. The results were compared to the number of active researchers in each subject area in order to determine which subjects had large or small web presences for their size.

The above shows examples of studies carried out using both web link and web pages classification exercises. However, Haas and Grams (1998) believed that classification systems for pages and links would be most effective if they were developed in tandem. They proposed an integrated classification system based on a content analysis of 75 web pages and the 1500 web links they contained.

This body of research, although by no means conclusive, shows that web-based studies should not only consider link counts, but also the motivations for linking, in order to ensure the validity of such research. However, the difficulty in both classifying link motivations and interpreting link counts appears to be a major problem for web link related studies and further motivation studies are vital for developing an understanding of how link counts should be interpreted.

2.3.2 Web Technology Analysis

Although most webometrics research to date has focused on hyperlinks, the reliance on commercial search engines to provide raw data, coupled with the success of Google (Brin & Page, 1998), has simulated interest in search engine performance and there is now a large body of research covering the quantitative aspects of search engine results (Thelwall, 2001e; Thelwall, 2002h; Upstill, Craswell & Hawking, 2003; Vaughan & Thelwall, 2004; Bar-Ilan, 2004d; 2005; Jansen & Spink, 2005; Bar-Ilan, Levene & Mat-Hassan, 2006; Bar-Ilan, Mat-Hassan & Levene, 2006).

Cronin (1984) assumed that more important or higher-quality journal articles would tend to be cited more and that citations often indicate that the work in the cited article has been built upon or otherwise used by the citing article. Google applies this assumption to the web, with its PageRank algorithm designed to find the most important pages on the web by analysing hyperlink structures. Google has now become so popular that company web site designers take into account how it analyses links when designing their sites' navigational structure (Park, 2002; Bar-Ilan, 2007). Interestingly, showing that the application of citation analysis techniques to the web is not always a one-way relationship, Google's success has led to PageRank's adoption for bibliometrics (Thelwall, 2002c; Thelwall & Vaughan, 2004a).

Since Almind and Ingwersen (1997) used the advanced Boolean features of AltaVista, this particular commercial search engine has been used extensively in webometric research, proving to be particularly useful for counts of pages over large areas of the web. Many early webometric studies, including Rousseau (1997), in which he retrieved all the occurrences of the search terms 'informetrics OR bibliometrics OR scientometrics', have used the Boolean tools available within the AltaVista search engine.

Although many webometric studies rely on commercial engines to gather link data, they have a number of drawbacks, not least that their crawling and reporting algorithms are commercial secrets and therefore not generally available to the research community. This presents a problem for their use in scientific research although researchers affiliated with commercial search engines have an option to use full search engine crawls. This approach has been used (e.g. Broder et al., 2000), although the crawls themselves have remained secret.

Bar-Ilan (1999; 2002a; 2002b) examined the performance of search engines over time, finding that they appear to lose information. Wouters, Hellsten and Leydesdorff (2004) confirmed this, adding that search engines are unreliable tools for scholarly research data collection, eroding the quality of information. Rousseau (2001) carried out a time series analysis of the number of hits in keyword searches over the year 1999, finding that search engines results at that time were too irregular to obtain reliable counts.

In addition to concerns about the accuracy and performance of commercial engines, a number of studies have expressed concern over the amount of the web, or even the publicly indexable web, commercial search engines actually cover. Lawrence and Giles (1998; 1999) calculated the coverage of different search engines on the web by monitoring queries. They discovered that the largest search engine at the time (NorthernLight) covered only approximately 16 percent of the web pages reachable and indexable by search engines. The results from Thelwall (2000), again using AltaVista, indicated that search engine coverage, even of large national domains, is extremely uneven and likely to lead to misleading calculations.

Snyder and Rosenbaum (1999) also observed large variations and inconsistencies, in particular concerning the AltaVista engine's link-page recovery. The irregularities of this engine have also been reported by Bar-Ilan (1999) in a longitudinal study as well as by Rousseau (1998/1999) who compared AltaVista and NorthernLight on a day-to-day basis over 21 weeks during 1999. The latter proposing that a method with which to combat irregular search engine results would be to use multiple search rounds and an averaging process. Rather than relying on the results from one search engine alone, Cronin et al., (1998) and Bar-Ilan (1998) used five and six (respectively) of the most popular search engines at the time in an attempt to compare and combine the results of their studies.

These are important findings for information scientists using search engines for their research and they should be aware of the problems that arise due to the inconsistent and unreliable nature of commercial search engines. If a commercial search engine is used in information science research, then this is a limitation which must be accepted and discussed during the study.

Despite these early disappointing results (Rousseau, 1998/1999; Snyder and Rosenbaum, 1999; Bar-Ilan, 1999), AltaVista subsequently became more stable (Thelwall, 2001c; Vaughan & Thelwall, 2003) and gave good coverage of academic web sites compared to a specialist web crawler (Thelwall, 2001a). However, the very latest webometric studies (Thelwall, 2007/8) advocate the use of the newly available MSN search engine tools as they appear to offer a comparable level of cover but allow a significantly greater number of queries to be returned.

Using a different approach, Kleinberg's (1999a) Hyperlink Induced Topic Search (HITS) algorithm uses a combination of page content and link structures to identify the most useful pages for the topic matching a search engine user's query. This is based on the assumption that the overall link structure of the web is not as important as that in the locality of the topic of concern.

2.3.3 Web Usage Analysis

This area of research is typically focused on log files of user's searching and browsing behaviour. An important application of server log analysis is to see which pages are most frequently viewed and to identify patterns of surfing with a view to improving site navigation (Huntington, Nicholas & Jamali, 2007). The principal drawback from a webometric perspective is that log files typically cover one site or all sites by the same server. Also, access to these logs is often restricted, and so these are generally not a good data source for studying multiple web site use. Pirolli and Pitkow (1999) analysed web site server logs over a ten day period to compare different path reconstructions and to investigate how past surfing behaviour predicts future surfing choices.

A second type of web usage analysis is to study the log files of commercial engines to analyse users' search sessions. A longitudinal study of Excite server logs (Spink et al., 2001) has shown that although search topics have shifted, there has been little change in user search behaviours even though there is increased use of advanced search features such as Boolean operators and relevance feedback. Jansen, Spink and Pederson (2005) concentrated on longitudinal transactional log analysis to examine the characteristics and changes in AltaVista use between 1998 and 2002.

2.3.4 Web Link Structure Analysis

Web link structure analysis is the main focus of the current study. The importance of web links is noted by the web's founder Berners-Lee (1997), who tells us that 'the web increases the power of our writings, making them accessible to huge numbers of people and allowing us to draw on any part of the global information base by a simple hypertext link'. Walker (2002), Galitsky and Levene (2002), Hawking et al., (2004) and Cothey (2006) argue that hyperlinks have become the 'currency of the web' and their studies concentrate on the 'value' and 'power' of links as enablers of access to knowledge on the web.

The majority of webometric link structure analyses have been carried out on academic web spaces, and research associated with this area comprises the following section of this literature review.

2.3.4.1 University Links (National)

A large number of link structure analyses have used national university links as an area of research. According to Middleton, McConnell and Davidson (1999), university web sites function as a tool for communication, providing access and promotion targeted at a variety of users, both internal and external. Using hyperlinks to investigate this aspect of online informal scholarly communication is now established as a practical and useful approach (Wilkinson, Thelwall & Li, 2003) and analysing the interlinking between universities within a single country presents an ideal scale for such a study. The number of objects to analyse (one site per university in a country) is manageable and counting all links to and from a university web site seems to give a sufficiently high level of aggregation to produce reliable results. The peculiarities associated with link creation and data collection seem to average out at the level of entire universities allowing trends to be more easily identified (Thelwall, Vaughan & Björneborn, 2005). Being a global document network initially developed for scholarly use (Berners-Lee & Cailliau, 1990), a considerable number of research articles have already been published concerning scholarly communication on the web (Björk & Turk, 2000; Zhang, 2001; Jacobs, 2001; Kling & Callahan, 2003; Roberts, 1999; Liu, 2004; Fry, 2006; Vasileiadou & Van Den Besselaar, 2006), again most originating in the hope that web links could be used to provide similar information to that extracted from journal citations. University link counts have also been suggested as a weak proxy for university research quality in countries where there are no comparative figures available (Thelwall et al., 2001).

Early university link studies investigated the relationship between interlinking counts of a selected set of universities on a national level and the research activity indicators of those universities. The first findings were negative (Smith, 1999; Thomas & Willett, 2000), and the reasons suggested included failures in the search engines used to obtain the link counts and the number of links created for reasons unrelated to research. Counts of links to a set of 25 UK universities were subsequently found to correlate significantly with their average research indicators using the five-yearly UK Government Research Assessment Exercise (RAE) (Thelwall, 2001d). This finding provided the first concrete evidence of a real association between research and links, although no claim was made that research caused link creation. Further positive correlations were found between research measures and inlink counts for UK universities, modelling the linking process using counts of links between pairs of 86 UK universities and regressing these against combinations of faculty size and average research indicators (Thelwall, 2002d). The university research figures used were estimates derived by multiplying RAE ratings of universities by total active researchers. The best predictor of total links between a pair of institutions was found to be the product of their two total research activity indicators and this gave evidence that outlink creation is not fundamentally different to inlink reception.

Comparable relationships were also found for Australia (Smith & Thelwall, 2002), Taiwan and Mainland China (Thelwall & Tang, 2003), Iran (Kousha & Horri, 2004) and Canada (Vaughan & Thelwall, 2005) using different measures of national research. Later studies found statistically significant correlations on a departmental level between UK computer science departments and their research ratings (Li et al., 2003) and between a higher education institution's investment in its academic library and its RAE rating (Oppenheim & Stuart, 2004). These successful studies provide evidence that hyperlinks bear some relationship to scholarly communication, although they were not necessarily caused directly by it.

In some of the studies mentioned above, when correlations were found they were not very high and graphs of the data highlighted individual universities which did not fit the pattern. In these instances, these anomalies were traced to individual cases where individuals or automated processes had created huge numbers of links. For example, a biochemistry database at Warwick University contained tens of thousands of links to a similar online database at the University of Cambridge, dwarfing the other link counts (Thelwall, 2002e).

It is important to bear in mind that although link counts can be shown to correlate strongly with research, university web link targets are very rarely academic papers, although many more links to academic papers are present in postscript and PDF documents on the web (Goodrum et al., 2001). Approximately 90% of links between national universities are created for reasons associated with educational and scholarly activity (Wilkinson et al., 2003), but even the most highly inlinked pages rarely contain high quality scholarly content (Thelwall, 2002i). A study of the 100 UK academic pages that were the target of most links from other UK universities found mainly general-purpose pages, such as university home pages and some departmental pages (Thelwall, 2003c).

Further investigations have unearthed specific relationships inherent in academic web spaces (Thelwall, 2002b; 2005; Payne & Thelwall, 2004). Another of these studies highlighted the apparent geographic grouping of UK academic institutions, finding that the extent of interlinking between pairs of UK universities decreased with geographic distance and that neighbouring institutions were more likely to interlink, particularly with respect to Scottish and Manchester universities (Thelwall, 2002f). This evidence of geographic clustering has since been reinforced by subsequent studies (Heimeriks & Van Den Besselaar, 2006) and shows that, despite the existence of various and numerous inter-university collaborations, the web is not divorced from physical reality.

Thelwall (2002d) demonstrated that counts of links to universities seem to be approximately proportional to the quadruple product of the size in academic staff numbers and research quality of the source and target institutions. This is consistent with the finding that universities with higher research activity indicators attract more links because they create more web pages, rather than because their individual pages are more likely to attract links (Thelwall & Harries, 2004a). This is a significant quantitative finding as, although it may appear that universities conducting more research attract significantly more links, in general, universities with better researchers attract more links because the researchers produce more web content, rather than because the content produced is of a higher quality. This is in contrast to the case for formal scholarly publications, where better scholars tend to produce articles that attract more citations (Borgman & Furner, 2002), and is a critical finding suggesting that link counts should not be regarded as a measure of research quality.

An investigation into personal home pages linking to UK universities found that this source of links gave very similar quantitative results to inter-university links, even though almost a third of the links were for recreational purposes (Thelwall & Harries, 2004b). It can be concluded that link attractiveness is relatively robust for UK university web sites, but that it is not exclusively dependant on the academic content of the pages linked to, even though link counts correlate highly with university research activity indicators. A further study, (Barjak, Li & Thelwall, 2007) showed only a weak relationship between a scientists recognition and homepage inlinks and, surprisingly, no relationship between research and inlink counts. Thoms and Thelwall (2005) also studied academic home pages, concluding that, whether the homepage is constructed by the academic or by the university, the identities of the individual are ultimately lost to the ‘governmentality’ of the university.

Other studies which have used university link analysis techniques to carry out research on a national level include Aguillo (2004) who used five major search engines to carry out a cybermetrics analysis of Indian universities, identifying that English is the predominant language of the Indian academic web space and that ten academic institutions are among the first hundred most visible Indian web sites and Bar-Ilan (2004c) who undertook a microscopic link analysis to categorise the interlinkage between 8 Israeli universities finding that these links were created to indicate reputation rather than research appreciation or collaboration. Pinto et al., (2005) carried out a quality assessment of 19 Spanish university's web sites focused on the European Research area, making proposals to improve the visibility of information on research while Kaliczynska (2005) took the statistical analysis of UK universities of Payne and Thelwall (2004) and applied the methods to Polish universities. Additionally, Vaughan and Thelwall (2005) found, by comparing different potential link attractors for Canadian universities, that faculty quality and language were the best predictors of inlink counts; French language universities in Canada attracted significantly fewer links than comparable English language universities. More recently, Onyancha and Ocholla (2007) carried out a web link analysis to compare Kenyan and South African universities according to several web-based indicators. They found that Kenyan universities, like most African universities, have embraced the Internet and its constructs fairly recently and hence most of their websites are at initial stages of construction. Comparatively, South African universities have made remarkable progress in their web presence, equalling counterparts in more developed countries.

The majority of university studies have analysed inter-university links in an attempt to identify inter-linking patterns and relationships, but some have focused instead on connections between universities and other sectors of society, such as commerce, industry and government (Leydesdorff & Curran, 2000; Thelwall, 2004a; Stuart & Thelwall, 2005; Stuart, Thelwall & Harries, 2006). The study by Chen et al., (1998) includes cross-national as well as cross-sectoral link connectivity studies highlighting cultural, political and technological interactions in the use of the academic web. Other studies concentrated on links between universities and libraries, both academic (Harpel-Burke, 2006) and public (Tang & Thelwall, 2005). The former found that 91% of medium-sized US universities had links to their library from their home page (seen to be an indication of the importance of the university library to the university webmasters), while the latter revealed that there is little interaction between public libraries in the US and US universities, highlighting the need for more collaboration between the two types of organisation.

Many previous web link studies have used commercial search engines for raw data but, although successful results have been obtained from them, their use is less than optimal due to concerns over their coverage, performance and reliability (Lawrence & Giles, 1998; 1999; Rousseau, 1998/1999; Snyder and Rosenbaum, 1999; Bar-Ilan, 1999; Thelwall, 2000). An alternative to the use of commercial engines in webometric research is to create a specialist information science web crawler, as called for by Bar-Ilan (2001) and Rousseau (2001), recognising the limitations in the use of commercial search engines for webometric research. Instead of having to rely on secondary data collected using commercial search engines, Thelwall (2001a; 2001b) developed and introduced an important methodological improvement for webometric investigations of academic web spaces in the shape of a specially designed web crawler (later known as SocSciBot and discussed in more detail in chapter 4.2) for collecting primary web data directly from the investigated academic web sites. This is not on a scale to rival commercial search engines, but it is capable of crawling all universities in a single country within a month. The raw data text files output by this crawler are collected as part of the University of Wolverhampton Academic Web Link Database Project (Thelwall, 2002/3) and have already been employed in many of the academic web links studies detailed here and are used extensively during this thesis.

However, although this information science crawler addressed many of the problems associated with the use of commercial engines, there are also a number of limitations inherent in the design of the data collection method. Web crawlers operate by following links and are limited in that they can only find pages that they are allowed to visit, are linked to or already know about and are linked to in a way in which the crawler can extract from the linking page. The number of pages found will depend upon the site, the crawler design and the parameters under which the crawler is operating.

A significant limitation in the information science web crawler used to collect the data for this study is its inability to crawl dynamically generated pages, as many universities have, over the period of this study, adopted technologies which integrate them into their web sites (although chapter 8 attempts to partially address this specific limitation). Dynamically generated pages are web pages that are created in response to web surfers' actions and do not exist before they are requested. The web page created is a genuine web page with its own unique URL but if a crawler visited the search engine site again, it would not find the same page as it was created in response to the query and then effectively destroyed. There are several web technologies that make it easy to do this, including PHP (Hypertext Pre-processor) and Microsoft's Active Server Pages (ASP). An academic web site using these kinds of technologies without the back up of HTML links would not be covered completely, although in some universities, including the University of Wolverhampton, certain types of pages such as main university and departmental sites have reverted back to standard links.

Another problem for web crawlers is presented by the use of obscured links. These are URLs that can be accessed by web users in ways that are difficult or impossible for web crawlers. The link extractor part of a crawler is not capable of extracting all links from web pages because some can be stored in formats that are in practice impossible for them to decode. For example, with programs running through the web browser, such as JavaScript, Java, Shockwave and Flash, it is not possible to easily extract URLs since these may be built by the code itself when running and embedding within the web page. Obscured links are an important threat to the validity of link analysis data and if any university sites in our data set used them extensively enough to prevent it being effectively indexed, then it may not be possible to conduct an effective analysis of the set. When using data from a crawler, its inability to find many types of dynamically generated pages and obscured links must be accepted as an unavoidable limitation.

These are important findings for information scientists using search engines or web crawlers for their research and they should be aware of the problems that arise due to the use of either. Concerns over the coverage, performance, reliability and availability of commercial search engines must be weighed against a web crawler's ability to not only locate relevant pages but also to then extract web links from them. However, certain studies show that, although the data collection methods used may vary the results may be similar (e.g., Bar-Ilan, 2004b and Wilkinson et al., 2003 carried out academic link classification exercises using a commercial engine and a specialist web crawler respectively, both reaching similar conclusions).

2.3.4.1.1 Web Impact Factor (WIF)

Many bibliometric studies have attempted to evaluate the impact of collections of journal articles by analysing their citation counts, and this led information scientists to investigate whether web link counts could be used as valid measures of online impact i.e. whether pages attract hyperlinks primarily for the quality or interest level of their content. If this were proved to be the case, hyperlink counts could be considered to measure some kind of online impact. Rodriguez i Gairin (1997) first introduced the concept of information impact on the Internet in a Spanish documentation journal and this was closely followed by Ingwersen's (1998) introduction of the concept of the Web Impact Factor (WIF) for national domains and individual web sites. This derived from obvious parallels with the Journal Impact Factor (JIF) published by the ISI for scientific journals receiving citations from scientific journals indexed in the ISI citation databases (Bollen et al., 2005; Garfield, 2006). However, the time periods for the WIF and the JIF are different (Noruzi, 2006b). The JIF measures citations made in journals published during one time period to articles published in another time period while the WIF is a 'snapshot' of a search engine database at a specific time. The WIF is therefore not the exact equivalent of the JIF although the WIF was inspired by the JIF.

The WIF is a metric designed to assess the average online impact of a set of web pages by counting the inlinking pages outside the set in question. Many different variants were proposed and tested but the most successful initially were the External Absolute WIF and the External Relative WIF. The former is simply the number of pages outside the web site (or TLD) being measured that contain a hyperlink to it while the latter is this figure divided by the number of pages inside the set. Both of these omit hyperlinks between pages within the target area, these often being for navigation purposes and therefore not useful indicators of external impact.

Ingwersen (1998) used the advanced search capabilities of the search engine AltaVista to count links to and from entire countries. However, the fluctuating performance of AltaVista at the time of the study yielded problematic variations in the calculated WIF measures. A further problem with the relative WIF in particular is that counts of pages within an area of the web can be substantially more unreliable than hyperlink counts. This is due to various factors including mirror site inclusion and design decisions about web page sizes and format (Thelwall, 2001d). In response to this, the WIF was later modified for universities by using another measure of university size, its full-time faculty (Thelwall, 2001d, 2003b), giving improved results. The modified (staff denominator) WIF counts links per full-time faculty member and hence avoids the problem of lower scores for institutions that publish large numbers of pages that are unlikely to be linked to.

The WIF has been used in several academic papers in order to find possible correlations to traditional research activity indicators (Thelwall, 2001d; 2002a; 2002d; Smith & Thelwall, 2001; 2002; Thelwall & Tang, 2003; Vaughan & Thelwall, 2005). The WIF has also made a significant contribution to academic hyperlink research by enabling deeper mining of university web link data through the factoring out of the research-related component (Li et al., 2003).

The correlations found between universities opened the door to attempts to mine deeper for patterns and to model the linking process. Metrics to measure the use of the web by universities (via outlinks) and the connectivity of a university with its peers were proposed (Thelwall, 2003b). These were termed the Web Use Factor (WUF) and the Web Connectivity Factor (WCF), and were based upon site outlinks and the links between pairs of sites respectively. The WCF was designed to restrict the impact of outliers on the data by assessing the link strength between a pair of universities to be the minimum of the count of links from the first to the second and the count of links from the second to the first. This was intended to remove unidirectional anomalies however, this more complex measure was not found to be significantly more robust, based upon the UK data set used (108 universities). An interesting development was the use of the measurements to provide baselines from which to compare the WIFs / WUFs / WCFs of individual universities in order to identify those that were not well connected on the web, perhaps indicating underlying problems in university web usage or publishing policies. WUFs were not found to be statistically less reliable than WIFs, despite being dependant upon the crawling of a single site to identify its outlinks, rather than upon multiple other sites to compile total inlinks. WUFs were also found to correlate strongly with average research statistics. Both measures are therefore supported by statistical evidence of their consistency at a general level, but both also show significant anomalies for individual web sites and are therefore not reliable for specific sites.

The WIF has proved to be a popular tool for other information scientists analysing web links. For example, the WIF has been used to compare the impact of web spaces with both on and off-line journals. Smith (1999) compares WIFs for the web spaces of Australian universities and Australian electronic journals, calling the WIF a 'useful measure of the overall influence of the web space', when specifically discussing universities and research institutions. For large organisations such as universities or research institutions, WIFs seem to be a useful measure of the overall influence of the web space however, for smaller measures such as electronic journals, the WIF is less reliable as a measure. An and Qui (2004) compared the JIFs of 42 Chinese engineering journals with external web link counts (from the Lycos search engine) and WIFs from corresponding journal web sites. The results showed that the correlation between the JIF and external link counts was borderline significant, while the correlation between the JIF and the WIF was statistically significant.

WIFs have also been used to analyse the web presence of Middle-Eastern countries TLDs. Noruzi (2006a) found that Middle-Eastern web sites (with the possible exceptions of Turkey, Israel and Iran) may achieve a lower visibility on the web because their language and culture are different from the current English-speaking mainstream of the web. Onyancha and Ocholla (2007) carried out a web link analysis to compare African universities according to several web-based indicators including the WIF. They found that Kenyan universities, like most African universities, have embraced the Internet and its constructs fairly recently and hence most of their websites are at initial stages of construction while South African universities have made remarkable progress in their web presence, equalling their counterparts in more developed countries.

Li (2003) carries out a review of the development and application of the WIF and more recent studies have focused on other web indicators, including the WIF, in an attempt to evaluate their relevance to current webometric research (Faba-Pérez et al., 2005; Katz & Cothey, 2006; Scharnhorst & Wouters, 2006)

2.3.4.1.2 Alternative Document Models

A fundamental definition for any web-related study is that of the web document, which should comprise a single indissoluble unit of coherent material (Cothey, Aguillo & Arroyo, 2006; Heimeriks, 2006). All early web link studies tended to use the web page as the primary source document for counting purposes until Thelwall (2002e) argued that this was not necessarily ideal and that other alternatives, specifically Alternative Document Models (initially known as Advanced Document Models) had the potential to produce better results. This is despite the fact that individual web pages are often the only choice if search engines are used for raw data, are the logical choice of primary web documentation and are by far the easiest basic web unit to identify and manage. Additionally, certain web-based research has to use commercial search engines to estimate hit counts and therefore cannot take advantage of ADMs.

The original ADMs exploited simple URL-based heuristics to automatically merge web pages for counting purposes (Thelwall, 2002e). ADMs are similar to an idea proposed by Björneborn (2001a) in that they are heuristics for grouping pages together into conceptual documents with the purpose of reducing the extent to which anomalies occur in web linking behaviour at the page level. They aggregate hyperlinks together based upon directories, domains and sites as well as the page and have been shown to be useful for circumventing anomalies in link data and conceptual problems with counting pages. In previous hyperlink studies, it was apparent that there were many cases in which one site contained thousands of hyperlinks to another, all created for essentially the same reason. This violated the implicit assumptions of hyperlink analysis in that each hyperlink should be of approximately the same importance as any other. An example of this type of problem, where individuals or automated processes have created huge numbers of outlinks, was shown in a study by Thelwall (2002e), where a Warwick University biochemistry database contained tens of thousands of links to a similar online database at the University of Cambridge, dwarfing the other link counts in the study.

The different levels of aggregation of pages and links were an attempt to remove the effect of spurious duplication of links which may occur when a website contains identical links on many pages. Using a directory or site ADM on such a case would dissolve the anomaly since the source and target pages would respectively be treated as a single 'macro-document'.

Using the terminology of Thelwall (2002e), there are four main ADMs in use, the page, directory, domain and site (or university) ADM:

Page: the page ADM is the default unit consisting of individual files on a web server. Each separate HTML file is treated as a document for the purposes of extracting links. Each unique link URL is treated as pointing to a separate document for the purposes of finding link targets. A web page in this context is identified with its URL. Any URL starting with http:// is allowed and URLs are truncated before any internal target designator symbol to avoid multiple links to different parts of the same page.

Directory: in the directory ADM, all pages in the same web server directory identified through the URL file name path are counted as one unit – as an aggregated 'macro document'. All HTML files in the same directory are treated as a document. All URLs are automatically shortened to the position of the last slash, and links from multiple pages in the same directory are combined and duplicates eliminated.

Domain: the domain ADM aggregates all pages with the same (subsite) domain name in their URL into a common unit of analysis. All HTML files with the same domain name are treated as a single document. This clusters together all pages hosted by a single subdomain of a university site. Domains are obtained by stripping any directory structure, file name, port number and password information from URLs, i.e. truncating each target URL just before the first slash it contained, if one was present.

Site: the site ADM (also called the university ADM) aggregates all derivative domain names into a single unit of analysis by specifying only the domain name ending. In the ADM framework, the site of University of Wolverhampton would function as a single unit of analysis embracing all pages from subsites, sub-subsites, etc., with URLs that contained domain names ending in '.wlv.ac.uk'.

A second approach using hybrid ADM counting methods was later developed, where multiple links to a target document from the same university were removed, giving even more statistically significant results (Thelwall & Wilkinson, 2003b). ADM metrics at different levels of aggregation using this approach effectively assess the range of the connection to each target university, although the outcomes are similar to the previous metrics. These 'range' models used a counting approach based on each of the four document models, but never counting the same link target more than once for the same source university. For example, if using the domain ADM and domains A and B in one university targeted domain C in another then this would score as only one link in the range model since the same domain was the target. The hypothesis behind the range models was that linking practices are frequently shared within a university and so multiple links to the same target could cause discrepancies in counts.

ADMs have been used extensively in academic web research as aggregated units of analysis. They were applied to a data set of UK universities and produced significant results, particularly for the directory and domain versions (Thelwall, 2002e). In this experiment, results for the four ADMs were compared with estimated research activity indicators for 108 UK university institutions. The underlying hypothesis was that a better ADM would generate results that would correlate more strongly with research-related figures. Although the Spearman correlation differences are small, the domain and directory graphs were much more linear, giving additional support for a closer relationship between links and research for the domain and directory ADMs, with the latter being the method of choice.

Thelwall and Aguillo (2003) utilised ADMs to undertake a health check of Spanish universities while Thelwall et al., (2003) calculated the percentage of the highest inlinked subject-based web sites of universities in Taiwan and Australia in 2003, classifying the sites found by crawling the university web sites in each country, applying the domain ADM to their link structures and then selecting the 100 highest inlinked subject-based web sites in each country. ADMs were also used to show significant correlation between university inlinks and research in Thailand, but the results were inconclusive for Mainland China, possibly due to scoring inadequacies (Thelwall & Tang, 2003) and to show that faculty quality and language are important predictors to links to Canadian university web sites (Vaughan & Thelwall, 2005).

ADMs have also been combined with a manual classification scheme. By restricting inlink counts to target pages mainly connected with research, the Spearman correlations reached a value of 0.949 for the directory range model, compared to 0.940 for the same counting model applied to all target pages and 0.920 for the standard file model applied to all target pages (Thelwall & Harries, 2003). The effect of applying the directory and domain ADMs overlapped with restricting the counts to classified pages, indicated that anomalies were predominantly found in pages not targeting academic content. This gives a high degree of confidence that links between university web sites are connected with scholarly activity in some way, despite the number that are created for recreational reasons, but still does not prove a cause-and-effect relationship.

It has been discovered that the domain and directory models were able to successfully reduce the impact of anomalous linking behaviour between pairs of websites, with the directory-based URL counting model being better for analysing interlinking between universities, at least in the UK (Thelwall & Wilkinson, 2003b; Payne & Thelwall 2004). However, the domain ADM is also a good choice, for example to compare links between individual UK, Australian and New Zealand universities (Thelwall, 2004b). Thelwall and Wilkinson (2003b) went on to state that the URL directory-based counting model appears to be a better model for analysing interlinking between universities than any of the standard models while Thelwall (2004b) found that the standard domain ADM emerged as the logical choice for comparison purposes when counting links from groups of large university web sites. Given the lack of a universal best choice of ADM, some studies seem to think that the best approach would be to use several (e.g., Thelwall & Vaughan, 2004a).

Thelwall and Vaughan (2004a) also introduced several new versions of the PageRank algorithm using ADMs, but with inconclusive results. The new approach seemed to work well for sets which included pages from other web sites but did not work well in ranking pages from the same site. Previous Google research had also used a link counting model equivalent to some ADMs (Bharat et al., 2001).

Adamic and Huberman (2000) have used a similar method to ADMs for aggregating web documents. They studied a crawl of 260,000 web sites, each representing a separate domain name and, by considering two sites to be connected if any of the pages at one site linked to any page in the other, they found that the distribution of inlinks between these sites followed a power law.

2.3.4.2 University Links (International)

Although many university web link studies have been carried out at a national level, as can be seen above, there have also been several international comparative studies of university web sites in other geographic areas including the European Union (Thelwall et al., 2002) and the Asia-Pacific region (Thelwall & Smith, 2002; Thelwall, 2004b).

Smith and Thelwall (2002) calculated WIFs for UK, Australian and New Zealand universities, and compared them to conventional measures of research output, i.e. rankings by Asiaweek magazine, the number of publications per staff member, and the number of citations per staff member. They found that, although there was a good correlation between the specially designed web crawler and AltaVista in estimating the link counts, the WIFs did not appear to correlate well with conventional measures of research output. They went on to compare linking between UK, Australian and New Zealand universities, and found that New Zealand was relatively isolated on the web, in line with findings from a previous bibliometric study for journals (Glänzel, 2001). A new measure was introduced; the propensity to link which was taken to be the total links from all universities in one country to all universities in the second, divided by the total number of faculty in the first country and also by the total number of faculty in the second country. A normalising calculation was used to allow these figures to be compared with links between universities in the same country. Not surprisingly, international links were dwarfed by national links, even for historically related and traditionally collaborating countries.

A larger follow-up study used AltaVista data and simple network diagrams (Thelwall, 2001c) to map the interlinking between universities in the Asia-Pacific region (Thelwall & Smith, 2002). In the four types of diagrams, the width of arrows between countries was proportional to raw link counts; links divided by target system size (total number of web pages in all universities in the target country); links divided by source system size; and links divided by source and target system size. All four diagrams were very different, but they showed that Australia and Japan were central to the academic link structure of the region, with smaller countries attracting attention disproportionate to their size. The diagrams were claimed to be most useful when comparing web collections of approximately the same size. Park and Thelwall (2006) examined the connectivity structure of links between university web sites in 25 Asian and European countries finding that university web sites in Asia were more heavily connected to European universities than linked to each other.

The hyperlink structure of the UK, Australian and New Zealand universities was also used by Thelwall and Wilkinson (2003c), finding strong scale-free regularities for page indegrees, outdegrees and connected component sizes, resulting in power laws similar to those previously identified for individual university web sites, and by Thelwall (2004b), in which this data was used to develop two methods with which to assess the reliability of link counts and to judge which of seven ADMs were most appropriate in each case.

There have been several published studies of European university web sites. Polanco et al., (2001) clustered universities throughout the European Union (EU) based upon the co-inlink (called ‘cositation’) analysis of 791 university sites from 15 European countries. A general survey of EU university web site sizes revealed a huge disparity between the West and East, one that could undermine attempts to use the Internet to integrate European research (Thelwall et al., 2002). AltaVista’s linguistic capability was used to examine the languages used for European web pages, and link pages in particular (Thelwall, Tang & Price, 2003). English was the major language for the whole of the EU with the exception of Greece, both for all pages and for internationally linking pages. Outside Greece and the English speaking nations, English language pages formed approximately half of all international link pages, with the other half taken up by indigenous languages. Not surprisingly, links between countries with a shared language were also common, particularly in the shared language. Apart from English, Swedish was the only other case of a language used extensively for international linking, mainly within Scandinavia, although Swedish is very close to other Scandinavian languages. Musgrove et al., (2003) successfully identified clusters of European countries based upon data for their universities interlinking patterns.

Thelwall (2002f) investigated alternative sources of links, classified by top-level domains. It was found that links to UK universities from .edu sources produced very similar results to those between UK universities, suggesting that academic link attractiveness is a phenomenon, numerically similar both internationally and nationally. Thelwall et al., (2007) assessed the web connectivity of international research groups finding that web connectivity seems to be particularly important for attracting overseas job applicants. Other current studies have also concentrated on research groups (Barjak & Thelwall, 2006; Thelwall, Barjak & Kretschmer, 2006) with the latter finding little evidence of gender differences between 9 European life sciences research groups. Cheng and Liu (2006) went on to classify the top 500 world universities into 21 types according to their disciplinary characteristics using a clustering method. The ranking of World Universities is also produced bi-annually by www.webometrics.info using webometric indicators to show the commitment of the institutions to web publication. If the web performance of an institution is below the expected position according to their academic excellence, university authorities should reconsider their web policy, promoting substantial increases in the volume and quality of their electronic publications. The ranking started in 2004 and is based on a combined indicator that takes into account both the volume of the web content and the visibility and impact of their web publications according to the number of external inlinks they receive.

2.3.4.3 University Links (Departmental)

In parallel with investigations into interlinking between universities, there have been several studies of interlinking on a departmental level. Success with research having a disciplinary focus could potentially lead to the development of techniques to measure subject communication patterns on the web, perhaps even giving early warning of emerging interdisciplinary trends. Thomas and Willett (2000) studied UK LIS departments, but found no significant correlation between inlink counts and research ratings. Significant differences between inlink counts and newspaper (US News) rankings were found later for US LIS schools (Chu, He & Thelwall, 2002), giving the first statistical evidence that departmental level studies could give information about scholarly communication. Since then, significant research and inlink count correlations have also been found for UK computer science departments (Li et al., 2003; 2005a; 2005b), mirroring previous university-level findings.

Significant correlations have also been established for US psychology and US chemistry departments (Tang & Thelwall, 2003; 2004). This research found that interlinking between US history departments was too low for patterns to be extracted and that there were significant disciplinary differences in patterns of interlinking within psychology, chemistry and history. This result supports the findings of Kling and McKim (2000) who stress the large differences between different scientific fields in the way electronic media, including the web, are implemented and utilised. No geographic trends were evident, perhaps either because geography is less important in the US than the UK or because the phenomenon is less evident within a single subject. In fact, history may actually be an anomaly in the humanities, because other humanities subjects seem to publish more on the web (Thelwall et al., 2003).

Chen et al., (1998) concentrated their attention on counting links between computer science department websites in 13 Scottish universities using pathfinder network diagrams and revealed a number of correlation relationships between structural connectivity measures and the organisational profile based on RAE ratings, teaching quality assessments, student-staff ratios and funding levels. Furthermore, linkage patterns from the 13 Scottish academic sites to commercial sites in UK and America highlighted the impact of culture and the appropriateness of information technologies on the acceptance of the web. This study was later criticised by Thelwall, Vaughan and Björneborn (2005) for not taking departmental size into sufficient account. The end results produced did reflect the profiles of the individual universities, but the survey was limited by the small sample size and a lack of variety in the institutions under study.

The above raises the important issue of scale when moving from analysing entire university sites to analysing individual departments. The studies reported indicate that inter-department link data are sufficiently strong in some disciplines to carry out an analysis, but the research relationships found have been much weaker than for whole universities, presumably due to the averaging effect of the larger units of analysis. There may be some research benefit in analysing the interlinking of departments in terms of cluster analysis approaches, but it is not yet clear that there is enough link data to do this meaningfully. An alternative approach may be for more text-based analyses to circumvent the problem of link sparseness, despite the greater technical difficulty with this mixed approach. In addition, the initial task of accurately identifying all web sites associated with a discipline is far from trivial (Li et al., 2003).

The scale issue is also a relevant factor when considering the kinds of institutional analyses common in scientometrics, i.e. evaluating all departments within a single university. The relatively small number of inlinks per department coupled with the inherently skewed nature of web linking means that inlink counts can only be very weak indicators of visibility or impact and should certainly be interpreted with great caution. However, the web is important to research communication and so departments attracting a low number of inlinks compared with other departments from the same discipline could reasonably be asked to at least address the issue of whether their web presence is effective or not (Thelwall, 2002g).

Other departmental-level research have sought reasons for inter-subject linking (Thelwall, Harries & Wilkinson, 2003) and assessed the visibility for UK university research areas, finding that Science and Engineering dominate the UK university web presence (Thelwall & Price, 2003).

2.4 Web Topology

2.4.1 Graph Theory

A graph is a mathematical representation of a network consisting of nodes (or vertices) connected by edges. In a directed graph the edges represent directional relations between the nodes whereas in an undirected graph the direction of the links is ignored. A branch of web-related study has sprung up which views the web as a collection of pages connected by links and considers it to be an example of a directed graph, with web pages corresponding to nodes and hyperlinks to edges.

Graph theoretic approaches have already been used to analyse citation networks in bibliometrics and scientometrics (Shepherd, Watters & Cai, 1990; Fang & Rousseau, 2001; Egghe & Rousseau, 2002) and hypertext research (Botafogo, Rivlin & Shneiderman, 1992) with great effect and so the web, as the largest, non-chaotic network for which topological information is currently available, is a natural area of study for information scientists and many attempts have been made to build topological models of the web through link analysis modelling efforts (Kleinberg & Lawrence, 2001; Thelwall, 2001c; Albert & Barabási, 2002) and physical layout modelling (Yook, Jeong & Barabási, 2002). Molyneux and Williams (2000) discuss, in some depth, the literature and history of Internet measurement.

Link analysis and graph theory have also combined successfully in other areas of research. Kleinberg's (1999a) HITS (Hypertext Induced Topic Selection) algorithm uses a combination of page content and link structures to identify the most useful pages for the topic matching a search engine user’s query. This is based on the assumption that the overall link structure of the web is not as important as that in the locality of the topic of concern. The HITS algorithm has been the subject of further analysis in conjunction with BibTechMon (Bibliometric Technology Monitoring) and the Internet Archive (Horlesberger & Scheibel, 2006; Ding et al., 2002) with some success.

Gibson, Kleinberg and Raghavan (1998), Flake et al., (2002) and Thelwall (2003a) concentrated on the idea of web communities, collections of pages within which each member page has more links within the community than outside the community. The primary argument being that, despite its decentralised and unorganised nature, the web self-organises to allow identification of highly related pages based solely on connectivity.

Following Adamic and Huberman's (2000) study of the power-law distribution of web links, Broder et al., (2000) validated the power law distributions for inlinks and outlinks (noting that Zipf-like distributions for inlinks give a better fit than the power law distribution) but also showing that graph theoretic methods can be used successfully to analyse structural aspects of the web. Using two full AltaVista crawls from May 1999 and October 1999 generating 200 million pages and 1.5 billion links, they built a database model of a web graph. This has come to be known as the 'Bow-Tie' model (Figure 2.1) and gives a valuable understanding of the intricate structure of the web. Over 90 percent of the links formed a huge connected group, provided the direction of the links was ignored. This central component splits into four roughly equal parts. The core is a Strongly Connected Component (SCC) in which all pages trace a directed link path to all others in the SCC. The other parts comprise a set of pages (OUT) that could be reached from the SCC by following directed links, a set of pages (IN) that could reach the SCC by following directed links, and the rest (TENDRILS). The remaining pages that were not connected in any way from the main 90 percent were dubbed DISCONNECTED. A problem with this study is that it is difficult to extrapolate to the whole web. AltaVista finds pages partly from user submissions of URLs, but mainly by following links from previously visited pages. As a result, pages that are not well linked to are more likely to be missed by its crawler. Thus DISCONNECTED is likely to be far greater for the whole web, but it is not possible to estimate how big it is because there is no practical way to automatically find pages that are not linked to.

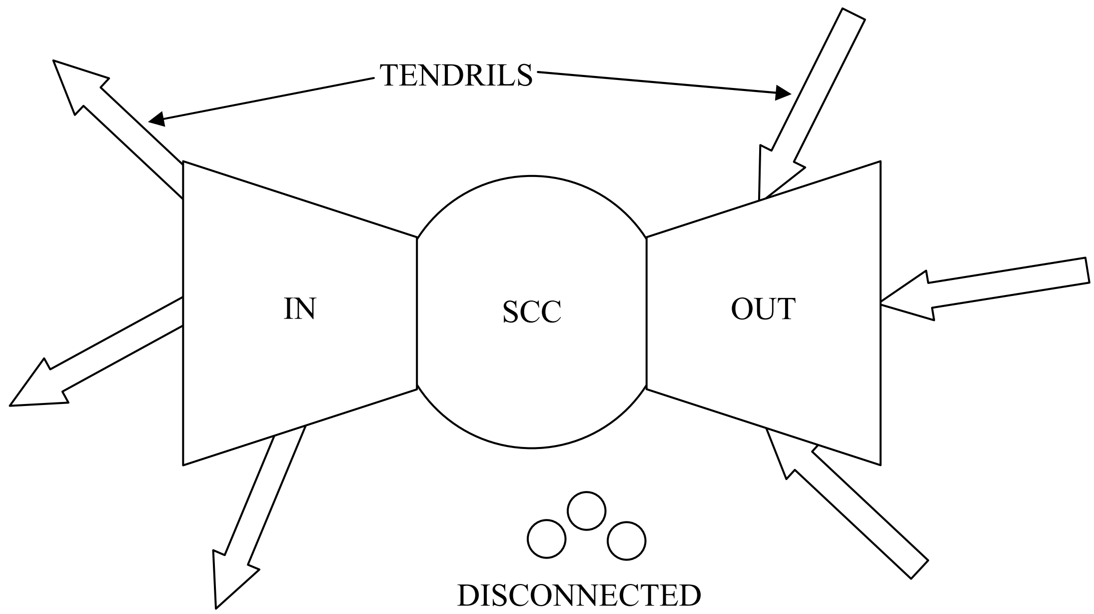

Figure 2.1 The 'Bow-Tie' Model

2.4.2 Power Laws

Although this sub-section falls under the web link structure analysis heading, it is worth noting that the research reviewed here was not conducted by dedicated webometricians.

The research of Huberman and Adamic (1999) identified a power law in the distribution of web pages. A power law occurs when the frequencies n of some variable x are proportional to $1/x^n$ and this meant that a diminishingly small proportion of sites had an increasingly large page count. This suggested that the web does not follow the mathematical models of random networks but instead exhibits the type of physical order found in nature, e.g. in magnetic fields or plant growth.

In a scale-free network (Barabási, Albert & Jeong, 1999; 2000), there is no typical node, i.e. there is no characteristic scale to the degree of connectivity. Scale-free distributions of web inlinks and outlinks exhibited long power-law tails, meaning that only a small number of web pages have many links, whereas the majority have few. Barabási and Albert (1999) argued that scale-free link distributions are rooted in two generic mechanisms: continuous growth and preferential attachment (the 'rich get richer' phenomenon, meaning that when new links are added to the web, they tend to link to pages which are already highly connected). In this framework, the web is seen as an open, self-organising system that grows by the continuous addition of new nodes and links where the likelihood of connecting to a node depends on the number of links already attached to the node (Dorogovtsev, Mendes & Samukhin, 2000). This condition is already well known in bibliometrics as the 'Matthew effect' and 'cumulative advantage'.

Indeed, other established bibliometric research techniques such as the use of Lotka, Zipf and Pareto power laws are also found in subsequent web studies

(Rousseau & Rousseau, 2000; Adamic & Huberman, 2002; Ajiferuke & Wolfram, 2004b). These laws are used either in a frequential or a ranked way. The frequential approach is the oldest one, as used by Lotka (1926) who showed that productivity is highly concentrated within a small number of researchers. Pareto's law is given in terms of the cumulative distribution function (CDF), i.e. the number of events larger than x is an inverse power of x. Pareto was interested in the distribution of income and he observed that there are a few multi-billionaires, but most people make only a modest income. The most characteristic example of a distribution per rank is Zipf's law. George Kingsley Zipf, a Harvard linguistics professor, examined how often words appeared in English texts. By ranking these numbers in a decreasing way, he observed that there was an inversely proportional connection between the presentation rank of a word and its apparition frequency. Zipf distributions have also been observed widely in computer science and are characterised by their long tails. Numerous works have shown equivalencies between the distributions per rank and the frequential distributions (e.g. Egghe, 1988) and the choice between one or the other depends on the study being undertaken.

Rousseau (1997) was the first information scientist to identify this kind of behaviour in the web. He used AltaVista to search for web pages containing the words 'bibliometrics OR informetrics OR scientometrics', retrieving 343 documents. The number of pages citing each of the pages in the results set was determined using AltaVista's link option. The study was able to fit appropriate Lotka functions to the data both for the number of retrieved pages per site, and for the number of citations to a site, identifying power-law distributions for the distribution of TLDs on a given topic. Huberman et al., (1998) showed that the surfing behaviour of web users follows Zipf-like distributions. The authors proposed a model of web surfing that explains the empirical findings on distributions of page hits observed at web sites.

Other early studies include Albert, Jeong and Barabási (1999) who, based on a subset of about 325,000 web pages, showed that both incoming and outgoing links obey appropriate power laws. Adamic and Huberman (2000) explained the distribution of the number of links to web sites using a power law while Adamic and Huberman (2001) reported finding almost identical power law distributions for number of site pages, visitors, inlinks and outlinks. The similarities can be explained by the fact that all four site characteristics evolve according to the same growth process.

Pennock et al., (2002) went on to explain why power law diagrams for web linking appear to have a highly skewed and hooked shape in terms of competing tendencies for new links to be allocated to existing highly linked pages, and for links to be allocated at random. In fact, the balance of competition varies by domain-specific types of pages. When the distribution of links was compared for the same type of page, it was found to be more log-normal than a power law, with additional variations by type. This shows that scale-free network growth models are not necessarily useful for explaining individual linking behaviour. Although the connectivity distribution over the entire web is close to a pure power law, the distribution within specific categories is typically unimodal on a log scale, with the location of the mode, and thus the extent of the 'rich get richer' phenomenon, varying across the differing categories. There seems to be a second linking law working in parallel with the power law, one that distributes links at random rather than predominantly to pages that are already highly linked. The revised model combines a power law with random linking and this made it easier to incorporate the possibility for new sites to gain high link counts. This model also explained the partial non-linearity of many power graphs, as a slightly hooked shape is evident, albeit to differing degrees, at the top left of each graph. The hooked shape in the graph indicates that this latter tendency is present, but is still dominated by preferential attachment.

The above research shows that power law models have become increasingly sophisticated over time, incorporating additional variables. For example, Adamic and Huberman (2002) have shown that site size, like site traffic, is highly skewed with a small number of sites responsible for a disproportionately large part of the web's volume and activity. In their model, the growth of the web is subject to two dynamics, the fact that the total number of sites is growing exponentially, and the fact that the fluctuations in the size of a particular site are proportional to the size of the site, 'the more pages a site has, the more likely it is that more pages will be added to it, just like the growth of a tree'. They hypothesised that the total size of the web will, like a tree, eventually become subject to resource limitations but that at the moment, the current web is still just a sapling, with plenty of potential for continued exponential growth.

There are many other examples of power law distributions to be found within the confines of the Internet, including the number of intermediate links along link paths between web pages (Albert, Jeong & Barabási, 1999). Barabási, Albert and Jeong (1999) showed that power laws apply to vertex connectivities in many large networks including the web. Faloutsos, Faloutsos and Faloutsos (1999) and Medina, Matta and Byers (2000) have identified corresponding power-law distributions which accurately characterise Internet topology while Chan et al., (2003) presented a graph layout algorithm in an attempt to simplify the visualisation of large power-law network topologies. Power laws, albeit with anomalies, have also been discovered within the graph structures of the national university publicly indexable webs from the UK, Australia and New Zealand (Thelwall & Wilkinson 2003c).

The web is a complex system which exhibits many characteristics found in many real world networks. The existence of a power law in the growth of the web not only implies the lack of any natural scale for the web, but also allows the expected number of sites of any given size to be determined without exhaustively crawling the web.

2.4.3 Small Worlds

Small-world theory stems from research in social network analysis (discussed in chapter 2.5), and has been popularised by the proposal of six degrees of separation concerning short distances between two arbitrary people through intermediate chains of acquaintances. Small-world web topologies are concerned with core information science issues such as navigability and accessibility of information across vast document networks. The emergence of small-world topologies on the web and in other evolving complex biological, technical and social networks can be attributed to the scale-free network features discussed above as, according to Albert and Barabási (2002), a heterogeneous scale-free topology is very efficient in bringing network nodes close to each other.

In a seminal paper, Watts and Strogatz (1998) introduced a small-world network model characterised by highly clustered nodes (as in regular graphs) yet with short characteristic path lengths between pairs of nodes (as in random graphs). They showed that in a small-world network it is sufficient for a very small percentage of 'long-range' links to function as short cuts connecting distant nodes of the network.

Albert, Jeong and Barabási (1999) conjectured that any pair of web pages would be connected by a short chain of links, with an average of only 19, but Broder et al., (2000) later showed that this was incorrect, as many pairs of web pages are not joined by chains of hyperlinks because many web pages are not connected at all. They examined around 200 million pages and 1.5 billion hyperlinks finding that more than 90 percent of the sample web pages form a single connected part if hyperlinks are treated as bi-directional. The probability that there was a hyperlink path between two randomly chosen web pages was only 24 percent. When there was a path, there was an average of approximately 16 hyperlinks in the path between pages. (The figure was much smaller, 6.83, for an undirected path).

The popularity of the small-world theory resulted in a large body of research in a wide range of scientific domains involving the web (Watts, 1999a; 1999b; Kleinberg, 1999b; 2000, Kleinberg et al., 1999; Barabási, 2001; Björneborn, 2001b; 2004; 2006). Although it is listed here as an aspect of web topology, many social science researchers also found that their work on small worlds within a social science framework translated effectively into a web-based environment.

2.5 Social Network Analysis

This section reviews hyperlink research conducted from the perspective of social network analysis within the topics of e-commerce, social movements, interpersonal, and international communication. The underlying belief is that the structural patterns of link connectivity can serve a particular social or communicative function.

In his highly-cited research, Granovetter (1973, 1983) argues that our acquaintances (weak ties) are less likely to be socially involved with one another than are our close friends (strong ties). The set of people made up of any individual and their acquaintances comprises a low density network (one in which many of the possible relational ties are absent) whereas the set consisting of the same individual and their close friends will be densely knit (many of the possible ties are present). In a famous experiment, Milgram (1967) wondered how long the average chain of acquaintances would be between pairs of random strangers. He selected 160 people from two towns in the US and gave them a letter, requesting that they forward it to somebody who they thought could help get it to the final recipient, a person in a distant US state. Of the 42 that reached their destination, the average number of intermediaries was a surprisingly short 5.5. This lead to a popularization of the notion of 'six degrees of separation' which is fundamental to the small-world theory discussed in chapter 2.4.2 above. Both of these examples have led to the successful application of social science techniques to Internet analyses. However, Elmer (2006) argues that much previous social network analysis has focussed too heavily on web hyperlinks as indicators of social networks and associations and that hyperlinks only suggest networking, i.e. they cannot specify the exact conditions and reasons for linking without additional forms of analysis.

Park and Thelwall (2003) reviewed and compared webometric information science approaches to those from social (or hyperlink) network analysis. They claimed that information science tended to emphasise data validation and the study of methodological issues, whereas social network analyses suggested how its existing theory could transfer to the web. Thelwall (2006a) outlines a theoretical framework for social science link analysis interpretation as a way to introduce information science style link analysis to a wider social science audience.

2.6 Longitudinal Research

Many information scientists have come to realise the importance of longitudinal studies with respect to the Internet while conducting their own research. According to Rousseau (1998/1999), ‘collecting time series should be an essential part of Internet research’. Guice (1998) argues that the Internet has its own history, even though most of the events connected with it took place just a few years ago, and that looking behind us more intelligently will make it easier to see paths ahead. Levene and Poulovassilis (2004) state that the web is highly dynamic in both the content and quality of the information it encompasses. In order to fully exploit its enormous potential as a global repository of information, they reason that we need to understand how its size, topology and content are evolving and this should then allow the development of new techniques for locating and retrieving information that would be better able to adapt and scale to the Internet’s change and growth. Kitchens and Mosley (2000) question the value of printed Internet guides as the web references they contain are too ephemeral while Thelwall and Payne (2005) echo this sentiment, highlighting a major problem endemic to web link analyses in that, as the web is continuously evolving, any web study may be out of date by the time it is published in the academic literature. Hence it is very important to know how all types of web link analysis results vary over time and a low rate of variation would lengthen the shelf-life of webometric results.

Harter and Kim (1996) were among the first to recognise the web’s ephemeral nature, finding that a third of electronic citations in e-journals were no longer available after a two year period. Koehler has produced a series of papers on web site constancy, permanence and persistence, in which he has analysed a random selection of 361 URLs since December 1996 (Koehler 1999a, 1999b, 2002 & 2004). This is believed to be the longest continuous study of a single set of URLs and, amongst the many longitudinal-based findings this body of research has produced, he has discovered that that the half-life of a web page is approximately two years and that web page content appears to have stabilised over time. He also found that different types of web pages (e.g., commercial, educational) behave differently. A limitation of the study is that these pages were randomly sampled from a single search engine (WebCrawler) in December 1996. These pages may not be an unbiased representation of web pages in general. An interesting trend, which has direct relevance to the current study, has emerged from his analyses; he has found that once a collection has sufficiently aged, it may stabilise in the sense that its URLs may become more durable in time. Koehler (2004) found that the collection of randomly collected URLs remained in a fairly steady state for two years after it lost approximately two-thirds of its population over a four year period.

Other longitudinal studies concentrating on identifying changes and trends in web pages include a study by Cho and Garcia-Molina (2003) during which they crawled a set of 720,000 pages on a daily basis over four months, and counted pages as having changed if their MD5 checksum changed. They found that 40% of all web pages in their set changed within a week, and 23% of those pages that fell into the .com domain changed daily. Fetterly et al., (2003) expanded on Cho and Garcia-Molina's study by crawling a set of 150,836,209 HTML pages once every week over an 11 week period. On completion of the crawl, they analysed the degree of change of each page, finding that pages drawn from servers in the .com domain changed substantially faster than those in other domains, while pages in the .gov domain changed substantially slower. Overall, they found that about 40% of all web pages changed within a week, and that it took about 50 days for half of all pages to have changed. Ntoulas, Cho and Olston (2004) measured the evolution of page content and link structure in 150 web sites over the course of one year finding a rapid turnover rate for web pages, although the rate of content shift of a given page was likely to remain consistent over time. That is, pages which change a great deal in one week will likely change by a similarly large degree in the following week. Conversely, pages that experience little change will continue to experience little change. Germain (2000) studied the accessibility of 64 URLs cited in 31 academic journal articles. After a three year period, almost 50 percent of the URL citations could not be accessed and two thirds of the journal articles contained corroded citations and this led her to question the use of URLs as citations for scholarly literature.

Brewington and Cybenko (2000), in an attempt to estimate the rate at which search engines must re-index the web in order to remain current, downloaded about 100,000 pages per day between March 1999 and November 1999. For pages that were downloaded six times or more, 56% did not change at all over the duration of the study, while 4% changed every single time. Ortega, Aguillo and Prieto (2006) carried out a longitudinal study of the state and evolution of 738 web sites in two different points in time (1997 and 2004). The main results confirmed a growth of web content and elements in the web, although a high degree of web content decay was also shown, with a claim that the web grows at the expense of the deletion of previous content. Bar-Yossef et al., (2004) concentrated on the issue of decay in the web over time, introducing a formal measure for decay and comparing it to the technique of counting the number of dead links on a page while Benbow (1998) found an attrition rate of 20% and 50% for web resources over two and three year periods respectively. Douglis et al., (1997) observed the rate of change of web pages in order to assess the benefits of caching (the less changes to the pages, the more useful it is) and found that content type and rate of access have a strong influence, while domain and size have little effect. Chi et al., (1998) presented new techniques for web ecology and evolution visualisation by using disk trees to represent a discrete time slice of the web ecology. A collection of these disk trees forms a time tube, which represents the evolution of the web over long periods of time.

Some of the significant longitudinal studies discussed above are compared explicitly in Table 2.1 below:

Table 2.1 Comparison of Web Page Longitudinal Studies

Study	Number of URLs	Origin of List	Type of URL	Type of Changes Sought
Koehler, 1999a	361	WebCrawler random URL generator	General	Web site and web page 'persistence'
Koehler, 1999b	361	WebCrawler random URL generator	General	Web site and web page 'constancy' and 'permanence'
Brewington and Cybenko, 2000	> 2 million	'The Informant' web clipping service	General	Changes to web URL and web page content
Koehler, 2002	361	WebCrawler random URL generator	General	The life-cycle of, and changes to, web page content
Cho and Garcia-Molina, 2003	720,000 (270 sites)	Simulated crawler on web page repository	General (popular)	Web page content (checksum) change
Fetterly et al., 2003	150,836,209	Mercator web crawler	General	Degree of change of each web page
Koehler, 2004	361	WebCrawler random URL generator	General	Web page 'persistence'
Bar-Yossef et al., 2004	1000	Yahoo! search engine	General	Web page 'decay'
Ntoulas, Cho and Olston, 2004	4.4 million (150 sites)	Google Directory	General (representative)	Link structure evolution and rate of creation of new content
Ortega, Aguillo and Price, 2006	145,092 (738 sites)	WebMapper (Microsoft Site Analyst)	Scientific	Web site growth and decay patterns

Bar-Ilan (1999; 2002a; 2002b) examined the performance of search engines over time, finding that they appear to lose information. For example, relevant URLs that were retrieved at a given time by a certain search engine were not retrieved by the same search engine at a later time, although they were known to exist and to be relevant. Wouters, Hellsten and Leydesdorff (2004) confirmed this, adding that search engines are unreliable tools for scholarly research data collection, eroding the quality of information. These are important findings for information scientists using search engines for their research and they should be aware of the problems that arise due to the inconsistent nature of search engines over time. Rousseau (2001) carried out a time series analysis of the number of hits in keyword searches over the year 1999, also finding that search engines results at that time were too irregular to obtain reliable counts. Bar-Ilan and Peritz (2004) studied changes which occurred to web documents related to the term 'informetric' over a five year period between 1998 and 2003, finding that pages were either completely static or changed often and considerably.

2.6.1 Sources of Time Series Data

Time series data can be obtained from various sources, including commercial search engines and the Internet Archive. For example, AltaVista allows searches for pages by their last modified (or creation) date (e.g., Leydesdorff & Curran, 2000). This approach results in partial data only as the search engine will ignore the actual creation date of pages, returning only the date of the last modification. The Internet Archive is becoming an increasingly important tool for conducting longitudinal studies of the Internet. It maintains a record of the evolution of the web, and is a key resource for webometricians. The particular differentiating feature of the Internet Archive is that although it operates like a commercial search engine, it keeps all retrieved copies of web pages so that changes in a page over time can be tracked and old pages that have been deleted from the web can still be found, allowing researchers access to old information. Longitudinal studies of the web can be conducted retrospectively using the Archive.

The Internet Archive Wayback Machine gives genuine time series data because it records all the dates on which a page was found, keeping copies of every indexed version so that changes can be tracked over time (e.g. Vaughan & Thelwall, 2003). The Archive can be used for an earliest known creation date for any page but this is not an accurate creation date because the earliest recorded date is merely when it was first found by the Archive crawler, and not when it was actually created. Another disadvantage of the Wayback Machine at the time of writing is that it does not allow global Boolean queries beyond simple wildcard matches, unlike AltaVista. An attractive alternative is to employ a crawler and collect data over time for a preset list of web pages or sites (e.g. Koehler, 2002), but this is not possible for retrospective studies. The potential importance of the Internet Archive for longitudinal and historical web research led to the need to evaluate its coverage. Thelwall and Vaughan (2004b) attempt to show whether or not there is an international bias in the Internet Archives coverage, and the results show that there are indeed large national differences in the Archives coverage of the web, with a bias towards US web sites. Although the bias is unintentional, researchers using the Archive need to be aware of this problem. Even if these issues are taken into consideration, the Internet Archive still represents a valid and reliable resource for information science research.

2.6.2 Search Engine Log File Longitudinal Research

Longitudinal analysis of the Excite search engine server logs has shown that search topics have shifted but there has been little change in user search behaviour (Spink et al., 2001) even though there was increased use of advanced search features such as Boolean operators and relevance feedback. Jansen, Spink and Pederson (2005) concentrated on longitudinal transactional log analysis to examine the characteristics and changes in AltaVista use between 1998 and 2002.

2.6.3 Other Studies

Other longitudinal research has been undertaken on specific web spaces including e-government web sites (Shi, 2006), higher education web sites (Hackett & Parmanto, 2005), health-related web sites (McMillan, 2001), news web sites (Vaughan & Dillon, 2000; 2006; Kutz & Herring, 2005) and corporate web sites (Yeung & Lu, 2004; Bjorn-Anderson & Elliot, 2005; Mendo & Fitzgerald, 2005; Heinze & Hu, 2006).

Several other Internet-related longitudinal studies are less relevant to this thesis. These include the Fomenkov et al., (2004) study of Inter traffic between 1998 and 2003, Ivory and Megraw's (2005) study of the evolution of web site design patterns from 2000 to 2003 and other studies concentrating on patterns in web-user search behaviour (Wang, Berry & Yang, 2003; Yuan, 1997; Cothey, 2002). Additionally, Huang, Chang and Chen (2006) carried out a purely bibliometric longitudinal study of Taiwanese universities from 1993 to 2003. Other bibliometric longitudinal studies include McCain and Salvuci's (2006) citation context analysis of Frederick Brooks' The Mythical Man-Month.

However, although it can be seen that there have been numerous academic hyperlink studies on departmental, national and international levels, very little research has been carried out from a purely longitudinal perspective. Björneborn (2004), in realising that the data set used in his study constituted a frozen snapshot of the publicly available university link structure data, specifically called for future longitudinal studies of academic web sites to be undertaken. Spink et al., (2001) also commented that many Internet studies are old, if not ancient, by the time they are published and so any study attempting to identify the amount by which the web changes over time would be of value to other web researchers, if only to give them an indication of how long their results are likely to remain current for.

2.7 Summary

As can be seen from the literature review above, much webometric research has originated from the belief that hyperlinks and citations share some common characteristics, and that applying established informetric methods and techniques to web data may uncover previously hidden patterns and relationships. Also, a large body of webometric research concerns itself with academic webs and, in particular, in identifying any correlations between university linking and research activity indicators. Academic webs seem to be a natural selection as an area of study for a number of reasons (identified in chapter 1.2) but not least that the original use of the Internet was intended for scholarly communication

Webometric analyses have now developed established methodologies and tools, and are used in encompassing research from other fields such as library science, computer science, theoretical physics and sociology. This is indicative of a widespread belief that links between web pages yield useful information and this belief is driven by the obvious importance of hyperlinks within the web and other high-profile uses such as Google.

However, although patterns can be extracted from hyperlinks, it is still the case that they are a largely unregulated and anarchic phenomenon. Great care must be taken to validate data when conducting hyperlink analyses to avoid drawing false conclusions because of data unreliability. Indeed, a sceptical researcher could claim that the problems critiqued in this chapter are so great that all web analyses lack value, but one response to this view, which is also a recurrent theme with critics of evaluative bibliometrics, is to demonstrate that web data correlates significantly with some non-web data (e.g. research activity indicators) in order to prove that the web data is not completely random. Nevertheless, the importance of links on the web is self-evident and, despite these problems, webometric research looks set to have a promising future.

The themes that emerge from link analysis research are valuable in a number of respects. Analysis of the link structure of the web suggests that the ongoing process of page creation and linkage, while very difficult to understand at a local level, results in a structure with a degree of order and may provide a basis for predicting the way in which the web will develop over time. The current research attempts to combine longitudinal study methods with university web link analysis in an attempt to identify and explain trends and patterns within academic web spaces over time.

3 Research Design

3.1 Introduction

This chapter discusses the primary research design used in this study, which includes the research philosophy and the research approach. It also defines, and discusses validity issues for, the research hypothesis.

3.2 Research Design

There are many research design methods currently in use including; experiments (common in pure scientific research); surveys (often used where large volumes of data are involved with quantitative methods of analysis); grounded theory (where the theory is generated by the observations rather than being decided before the study); ethnography (a phenomenological methodology which uses observed patterns of human activity); action research (where the research takes more of the form of a field experiment); modelling (where particular models are developed as the focus of the research activity); operational research (which looks at activities and seeks to understand their relationship, often with particular emphasis on operational efficiency) and case studies (which seek to understand social phenomena within a particular setting).

However, all of these research design methods stem predominantly from one of three main types of research design (Sridhar, 2007); exploratory research, descriptive (statistical) research and experimental design (hypothesis testing).

Exploratory research is often conducted because a problem has not been clearly defined, or its scope is unclear. It allows researchers to familiarise themselves with the problem or concept to be studied, and perhaps generate hypotheses to be tested. It is the initial research carried out, prior to more conclusive research being undertaken, and can help determine the best research design and/or data collection methods. Exploratory Research can be quite informal, relying on secondary research such as reviewing available literature, or qualitative approaches such as discussions, interviews, focus groups, case studies or pilot studies. This means that although the results of exploratory research are not usually useful for decision making by themselves, they can provide significant insight into a given situation.

Although there is an element of exploratory research in the current study, previous research has gone some way to defining the problem of a lack of longitudinal study within academic spaces and so this research is predominantly quantitative in nature and relies on a more descriptive research design, as defined below:

Descriptive (statistical) research has the purpose of providing information that is useful in reaching conclusions or decision-making. It tends to be quantitative in nature and relies on both primary research (data collected specifically for the study at hand) and secondary data (a review of all related literature or research in order to determine whether any data sources exist already that can be brought to bear on the problem at hand). It provides data about the population being studied and is used when the objective is to provide a systematic description that is as factual and accurate as possible. One of its major limitations is that it cannot help determine what causes a specific behaviour, motivation or occurrence (identified, and discussed, in chapter 2.3.1). In other words, it cannot establish a causal research relationship between variables and this is a theme of the current research. While a correlation between academic research indicators and counts of links is demonstrated on a number of occasions, it is noted that no causation is ever established.

Two commonly used types of descriptive research designs (Wynar, 1971) are:

- Observation. This is a primary method of collecting data by human, mechanical, electrical or electronic means. The data being collected can concern an event or other occurrence rather than people and observational techniques can be part of qualitative, as well as quantitative, research techniques.
- Survey. The survey technique involves the collection of primary data about subjects, usually by selecting a representative sample of the population under study through the use of a questionnaire. It is a very popular since many different types of information can be collected, including attitudinal, motivational, behavioural and perceptive aspects.

Experimental design is the basic design of logical proof. It helps in testing hypotheses, leads to inference on causality and requires controlled observation of change and development in variables. In other words, experimental or hypothesis testing studies involve studying causal relationship between variables and draw inferences about causality. Causal research is undertaken if the objective is to determine which variable might be causing certain behaviour, i.e. whether there is a cause and effect relationship between variables. In order to determine causality, it is important to hold the variable that is assumed to cause the change in the other variables constant and then measure the changes in the other variables. This type of research is very complex and the researcher can never be completely certain that there are not other factors influencing the causal relationship.

There are two research methods for exploring the cause and effect relationship between variables:

- Experimentation. This highly controlled method allows the researcher to manipulate a specific independent variable in order to determine what effect this manipulation would have on other dependent variables. (Most empirical studies in social and information studies are quasi-experimental i.e., groups have been constituted by means other than random selection).
- Simulation. A sophisticated set of mathematical formula is used to simulate or imitate a real life situation. By changing one variable in the equation, it is possible to determine the effect on the other variables in the equation.

In summary, the overall research design for the current study is one based on a descriptive (statistical) research approach, employing mainly observational techniques. There is also a secondary aspect of experimental design, as some causal research is carried out during the link classification exercise in chapter 7, but this is fairly weak.

3.2.1 Research Philosophy

For the current study, selecting an overall research philosophy was the choice between two primary alternatives; positivist or phenomenological. The following table (Easterby-Smith, Thorpe & Lowe, 1991) highlights some of the main issues for each philosophy:

Table 3.1 Comparison of Positivist and Phenomenological Philosophies

	Positivist paradigm	Phenomenological paradigm
Basic beliefs	The world is external and objective Observer is independent Science is value-free	The world is socially constructed and subjective Observer is part of what observed Science is driven by human interests
Researcher should	Focus on facts Look for causality and fundamental laws Reduce phenomenon to simplest elements Formulate hypotheses and then test them	Focus on meanings Try to understand what is happening Look at the totality of each situation Develop ideas through induction from data
Preferred methods include	Operationalising concepts so that they can be measured Taking large samples	Using multiple methods to establish different views of phenomena Small samples investigated in depth or over time

Basically, the positivism paradigm assumes that human behaviour is determined by external stimuli and that it is possible to use the principles and methods traditionally employed by the natural scientist to observe and measure social phenomena. The phenomenological paradigm is a research methodology which has its roots in philosophy and which focuses on the lived experience of individuals

Given the research aims and objectives as outlined in chapter 1.2, this research follows a predominantly positivist paradigm. This was done recognising that the following parameters identified by Hussey and Hussey (1997) for a positivist paradigm apply to the current study.

- It tends to produce quantitative data.
- Data is rich and objective: the quantitative data would be rich by nature, and the data collection methods employed (chapter 4) should be objective.

- Data reliability is high: the data is consistent over time and the results of the study can be reproduced using similar methodologies.
- Data validity is high: data validity should be high, mainly due to empirical nature of the data collection methods.

However, phenomenological aspects of this thesis include motivation analysis and correlation tests as multiple perspectives on the issue of links verses research activity indicators.

3.2.2 Research Approach

Research can have elements which are based upon a non-empirical approach, an empirical approach, or a combination of both. However, the research approach used in the current study is primarily empirical. The American Heritage Dictionary of the English Language (Fourth Edition, 2000) defines empirical research as:

- Relying on or derived from observation or experiment
- Verifiable or provable by means of observation or experiment
- Guided primarily by practical experience and not theory

This definition tells us that empirical research is research which is based on observed and measured phenomena and is research that derives knowledge from actual experience rather than from theory or belief.

For an empirical approach, the three primary dimensions which can be evaluated for use are given below. These do not necessarily represent a simple either/or choice, and the extent to which each element of the approach applies to the current research is considered:

3.2.2.1 Qualitative / Quantitative

Examples of quantitative methods include survey methods, laboratory experiments and numerical methods such as mathematical modelling. Examples of qualitative methods are action research, case study research and ethnography.

This study, concerned with identifying trends and patterns in the link structure of academic webs by performing data analysis on database text files is strongly biased towards using quantitative methods although a qualitative aspect is used during the link classification exercise in chapter 7.

3.2.2.2 Deductive / Inductive

Hussey and Hussey (1997) define deductive research as ‘a study in which a conceptual and theoretical structure is developed which is then tested by empirical observation; thus particular instances are deducted from general influences.’ Deductive research is a study in which theory is tested by empirical observation. The deductive method is referred to as moving from the general to the particular.

Inductive research is a study in which theory is ‘developed from the observation of empirical reality; thus general inferences are induced from particular instances, which is the reverse of the deductive method’ (Hussey and Hussey, 1997).

Many researchers advocate taking a middle-ground, striking a balance between pure induction (theory-building) and pure deduction (theory-testing) and the current study reflects this. It is mainly inductive, as it involves moving from individual observation to statements of general patterns or trends found within the academic web space data, however, there is also a deductive element in drawing upon bibliometric theories and testing them with academic web data.

3.2.2.3 Subjective / Objective

Another significant choice which exists in the research approach to be adopted is the extent to which the researcher is subjective (involved in or has an influence on the research outcome) or objective (distanced from or independent) in the execution of the fieldwork (empirical work).

The traditional assumption in science is that the researcher must maintain complete independence if there is to be any validity in the results produced. In the current study, the positivist, empirical approach used is, by its nature, objective.

In summary, the overall methodology is one based on a descriptive (statistical) research approach. It adopts a mainly positivist research philosophy and an empirical research approach; is objective rather than subjective (having little involvement by the researcher); is inductive in terms of theory building; and uses mainly quantitative methods.

This study also employs an element of grounded theory (where the theory is generated by the data, rather than the other way around) although the methodology is not rigorously followed. The study also includes several examples of modelling where models (including linear regression models) are developed as the focus of the research activity.

Data collection methods are dealt with in chapter 4.

3.3 Research Hypothesis

Given the research questions in chapter 1.3, the research hypothesis is defined as a statement of expected results. A hypothesis is more specific than a research question, and aims to provide clarity and to formalise the research question. A hypothesis can also assist in determining the direction a study will take. The literature review in chapter 2 underpins the following research hypothesis (H1).

Significant, identifiable changes in the hyperlink structure of the academic web spaces of the UK, Australia and New Zealand have occurred over the last six years.

There are two terminologies which are used for defining a hypothesis. They are variable and construct. A variable is defined as a general class of objects, events, situations, characteristics and attributes that are of interest to a study. A construct is defined as the concept (or idea of phenomena) that has to be defined (or specified), so that a variable can be measured (or quantified) on the basis of the construct.

Research hypotheses are the specific testable predictions made about the independent and dependent variables in the study. In this case, the dependent variables are the academic hyperlinks of UK, Australian and New Zealand universities, and the independent variable is time. Time is what causes the dependent variables to change and so this study focuses on the dependent variables (UK, Australian and New Zealand university hyperlinks) to see how they responds to the change made to the independent variable (time).

In order to define the research hypothesis, the construct which underlies the hypothesis is to be identified. This thesis uses 'the changes in the hyperlink structure of the academic web spaces of the UK, Australia and New Zealand' as a construct.

It is important to asses the validity of the hypothesis and the following paragraphs discuss the types of validation (Cook & Campbell, 1979) carried out.

Statistical Conclusion Validity: This asks 'is there a relationship between the two variables?' and, in this study the question might be worded 'is there a relationship between the hyperlink structure of UK, Australian and New Zealand universities and time?' At this stage, we hypothesise that there is such a relationship, and this research will attempt to identify changes within the relationship.

Internal Validity: This describes the ability to show that the independent variable was responsible for the change in the dependent variable, i.e. assuming that there is a relationship in this study, is the relationship a causal one? Much previous webometric research shows correlation between university link structures and other variables such as site size or research activity measures. However, many also point out that these correlations are not in themselves evidence of causation as both variables may be influenced by another factor. This study does hypothesise that the hyperlink structure of UK, Australian and New Zealand universities will change over time (and due to time) and some effort will be made to reduce the effect of influencing factors through a process of normalisation where possible.

Construct Validity: Assuming that there is a causal relationship in this study, the most critical point in the construct validation is to ensure that given a hypothesis, what is measured is what was intended to be measured. Failing to ensure construct validity can result in incorrect outcomes, which in turn, can lead to the wrong conclusions. In other terms, were the ideas of the cause and the effect operationalised well? To ensure the construct validity of this study, the raw data and results produced by the SocSciBot crawler and database were compared to similar studies where possible, and all output from the SocSciBot tools and bespoke programs underwent manual checking using realistic test data.

External Validity: Assuming that there is a causal relationship in this study between the constructs of cause and effect, could this effect be generalised to other places or times? This study claims that its research findings will have implications over different time periods and for other countries' academic webs. One key to having external validity is to have a large, representative sample of subjects and the university hyperlink structures of the UK, Australian and New Zealand academic web spaces seem to be a natural choice for webometric research as they have similar economic, linguistic and cultural backgrounds. The current study hopes to show that its results are consistent over time, and across different countries' academic web spaces.

3.4 Summary

In defining and examining the different types of research methodologies, philosophies and approaches, it can be seen that, for the current study, the overall research design is one based primarily on a descriptive (statistical) research approach, employing mainly observational techniques. It adopts a mainly positivist research philosophy and an empirical research approach; is objective rather than subjective (having little involvement by the researcher); is inductive in terms of theory building and uses mainly quantitative methods. This study also employs elements of modelling, exploratory research and grounded theory (in that the theory is generated by the data, rather than the other way around).

The overall research design used during the current study should be capable of supporting the research hypothesis defined and validated (using conclusion, internal, construct and external validity, to varying degrees) in chapter 3.3 and should enable the research questions stated in chapters 5 – 9 to be answered conclusively.

Chapter 5 employs a quantitative, mathematical modelling approach to answer the research question of whether the relationship between academic web and research activity indicators is best modelled by a linear trend or a power law. In doing so, this chapter validates previous assumptions and lays the foundation for future research. A descriptive (statistical) research design was essential in answering the research questions outlined in chapter 6 and the purely empirical approach, using correlation testing, showed that consistently higher results are obtained when the web links are aggregated at the directory and domain level.

Chapter 7 addresses the research question of how and why the distribution of types of academic web links changes over time and considers whether university web sites publish the same kind of information and use the same kind of hyperlinks year on year. This chapter is a first attempt to combine a link classification exercise with a longitudinal study and, in doing so, employs different research techniques in that, for this chapter alone, the research is of a more qualitative nature, using a phenomenological philosophy and an experimental design.

Chapters 8 and 9 employ similar research designs in order to answer research questions dealing with changes in trends for inlinks, outlinks and site size for academic institutions i.e. primarily a descriptive (statistical) research approach, employing mainly observational techniques, adopting a positivist research philosophy and an empirical research approach. These studies also extensively employ quantitative elements of mathematical modelling, both graphically and using linear regression techniques with correlation testing.

4 Research Methods

4.1 Introduction

This chapter answers the following two main questions:

1. How was the data collected?
2. How was it analysed?

Rousseau (2001) states that 'most cybermetrics research results more in statements of principle (showing what could be possible if one would be able to collect reliable data) than in exact results'. This chapter details the various data collecting methods used during the course of this research. It is important to show the research methods used for data collection, as the method affects the results. Knowledge of the data collection methods aids in the evaluation of the validity and reliability of the results and the conclusions drawn from them.

The data for the current study was collected in a way that is designed to be consistent with accepted webometric practices and are appropriate to the objectives of the study (chapter 1.2). The methodology also discusses the problems which occurred and the ways their impact was minimised. It is recognised that it may be useful for other researchers to adapt or replicate this research, and so the reasons why a particular method was chosen are given, and sufficient information is given, in this chapter and the annexes, to allow others to reproduce the work.

4.2 SocSciBot Crawler and Tools

It may be possible for an information science researcher to find all the web pages required for a small-scale webometric study by browsing alone. However, this approach is impractical for large-scale research such as the current study, where a commercial search engine or web crawler would be more appropriate. A web crawler, also known as a web spider or robot, is a piece of software that can automatically and iteratively download pages and extract their links.

Rousseau (2001) and Bar-Ilan (2001), in recognising the limitations in the use of commercial search engines for webometric research, both called for the construction of a cybermetric / scientometric web search and data collection tool. In partial response to this, Thelwall (2001a) created a specialist information science web crawler (later known as SocSciBot) for academic web sites.

This crawler crawls all HTML pages on an academic web site by following links, typically starting at the target university's home page and following links to the same site iteratively until all known pages have been visited. On university sites that have no embedded links on the home page (using for example a pull-down menu for selecting pages), an alternative web page was selected as the starting point for the crawl, usually a page with a list of links to the department home pages at the university.

There are a number of known limitations associated with this approach. Crawlers find new pages by following links and so are likely to miss many isolated pages and collections of pages. Pages that are linked to can also be missed if the links are in a form that the crawler does not understand (e.g. JavaScript), if the pages are password-protected, if the server is temporarily down or if the owner requests that the page is not to be crawled (e.g. through the robots exclusion protocol using the *robots.txt* convention). A web site may contain many web pages with different URLs but identical content and, in this case, such duplicate pages were identified and automatically excluded from the original link data set. As can be seen from the research, the SocSciBot crawler covers websites accurately in the sense of comprehensively testing for and eliminating duplicates but the results cannot claim to be complete as crawler coverage will typically be less than 100 percent of a site.

The web crawler also excluded mirror sites, online e-journals, and hosted web sites of external organisations as they were considered not to represent content created at the host institution. Furthermore, anomalies such as web pages containing thousands of automatically generated links were excluded because they would bias inter-university link counts. Although the choice of universities selected for the crawl was based on formal university listings (e.g. the Times Higher Education Supplement, which includes almost all official universities as well as most of the largest non-university HE institutions for the UK), manual intervention was needed in order to identify omissions and undesirable web pages and subsites (Thelwall, 2001b), although this process is error-prone due to its reliance upon human involvement. Note also that since 2001 a number of new universities have been created.

The following list gives a summary of some of the more important issues dealt with by the crawler:

- Duplicate pages within the same university are rejected
- Mirror sites, when identified, are rejected
- Subsites with derivative domain names are included (e.g. www.scit.wlv.ac.uk was included as part of www.wlv.ac.uk)
- Sites without HTML links on their home page are crawled from an alternative starting point, such as a departmental home pages list
- Pages with URLs containing a question mark are not crawled.

The output link structure of the crawler consists of a separate text file for each university, giving a list of the URLs of all source pages crawled together with all identified target URLs referred to in the page, with duplicate URLs removed and all URLs truncated at the first '#' character. This last point means that in one page there cannot be links to two or more parts of a common target page.

4.3 University of Wolverhampton Academic Web Link Database Project

The university text files produced by the SocSciBot crawler are freely available for public use (cybermetrics.wlv.ac.uk/database) through the on-going University of Wolverhampton Academic Web Link Database Project (Thelwall, 2002/3). This project has been collecting national university web link data since 2000 and now includes link data from UK, Australian, New Zealand, Chinese, Taiwanese and Spanish university web sites. The original purpose of the databases was to provide link data for WIF investigations (Ingwersen, 1998) concerned with identifying correlations between link counts between university sites and their research activity measures (e.g., Thelwall, 2001a; 2001d; 2001e; Smith & Thelwall, 2001; 2002).

While this project has only been collecting university link data since 2000, in the context of web analysis this is a long-term perspective and has already been used to provide significant insight into the patterns and relationships inherent in academic hyperlinks (Smith & Thelwall, 2002; Li et al., 2003; Thelwall & Harries, 2004a).

This study uses data for the universities of the UK, Australia and New Zealand over a six year period beginning in July 2000. Table 4.1 below shows the database number, country and dates on which the crawl took place. Although the crawls were not taken at exactly the same time period each year, attempts were made, especially from 2001 onwards, to crawl each national academic web space every year, and at roughly the same time each year. Given the similarities between each academic web space, the slight time discrepancy seems unlikely to significantly affect the results.

Table 4.1 University of Wolverhampton Academic Web Link Database Project Numbers, Countries and Crawl Dates

Database Number	Country	Crawl Dates
1	Australia	July-August 2000
2	New Zealand	July-August 2000
3	UK	June-July 2000
4	UK	July 2001
5	Australia	October 2001-January 2002
6	New Zealand	January 2002-February 2002
9	UK	June-July 2002
11	New Zealand	January 2003
12	Australia	February-March 2003
13	UK	June 2003
14	New Zealand	December 2003
15	Australia	February 2004
16	UK	June 2004
18	New Zealand	January 2005
19	Australia	January-March 2005
20	UK	July 2005
21	New Zealand	January 2006
22	Australia	April 2006

4.4 Bespoke Tools

Although a suite of tools is available online (socscibot.wlv.ac.uk) to process the SocSciBot crawler text files, these were not sufficient to produce all of the longitudinal results required in the current study. The results required necessitated the production of bespoke, special-to-type programs written to analyse the raw data.

These bespoke tools were produced using Visual Basic 6 and invariably sorted, counted and filtered the data based on user-defined criteria. Various incarnations of the programs were used during the course of this research, dependant on the outcome required, but including the removal of duplicate data, aggregation (using ADMs), the removal of self-links and the comparison of inlinks and outlinks. These programs are freely available from the author on request.

4.5 Internet Archive

As noted in chapter 2.7, an increasingly important tool for conducting longitudinal studies of the Internet is the Internet Archive. This maintains a record of the evolution of the web, and is a key resource for webometricians. The particular differentiating feature of the Internet Archive is that although it operates in some ways like a commercial search engine, it keeps all retrieved copies of web pages so that changes in a page over time can be tracked and old pages that have been deleted from the web can still be found. This allows researchers access to old information and means that longitudinal studies of the web can be conducted retrospectively via the archive. The Archive thus provides means for so-called 'web archaeology' (Björneborn & Ingwersen, 2001) for the retrieval and verification of links, web pages and web sites that otherwise may have disappeared from the dynamic web.

Björneborn (2004) found that even if the Internet Archive does not cover the entire web, over 90% of the investigated UK academic subsites had top home pages indexed in the Archive, concluding that the Internet Archive is thus an excellent web archaeological tool, at least for investigating the UK academic web space.

This research has employed the Internet Archive's Wayback Machine to examine how and when university web site resources have changed over time. For example, it can be used to identify links to pages the use of which is not immediately apparent from their URLs, and to ascertain when those pages were moved or withdrawn.

4.6 RAE Ratings

Much early webometric research concentrated on identifying correlations between links to and from a university web site and research activity measures. For the UK, this measure is based on the Research Assessment Exercise (RAE) which is a peer review, subject-based process that assesses the quality of research in UK universities and colleges to enable the four higher education funding bodies (England, Scotland, Wales and Northern Ireland) to distribute public funds for research selectively on the basis of research quality ratings. The institutions conducting high-quality research receive a higher proportion of funding.

The first RAE was undertaken in 1986. For the first time it introduced an explicit and formalised assessment process of the quality of research. Further exercises held in 1989, 1992 and 1996 became gradually more transparent, comprehensive and systematic. The fifth (and last) exercise in 2001 was the most rigorous and thorough exercise to date. It considered the work of almost 50,000 researchers in 2,598 submissions from 173 Higher Education Institutions and around £5 billion of research funds were distributed in response to it. All of the research carried out in this thesis dealing with UK research indicators uses this data. The next RAE is due to be carried out in 2008.

The RAE provides quality ratings for research across all disciplines using a standard scale ranging from 1 to 5, although the use of sub-categories results in the seven grades shown below. (Grades are determined by how much of the work is judged by a peer panel to reach national or international levels of excellence).

5* Quality that equates to attainable levels of international excellence in more than half of the research activity submitted and attainable levels of national excellence in the remainder.
5 Quality that equates to attainable levels of international excellence in up to half of the research activity submitted and to attainable levels of national excellence in virtually all of the remainder.
4 Quality that equates to attainable levels of national excellence in virtually all of the research activity submitted, showing some evidence of international excellence.
3a Quality that equates to attainable levels of national excellence in over two thirds of the research activity submitted, possibly showing evidence of international excellence.
3b Quality that equates to attainable levels of national excellence in more than half of the research activity submitted.
2 Quality that equates to attainable levels of national excellence in up to half of the research activity submitted.
1 Quality that equates to attainable levels of national excellence in none, or virtually none, of the research activity submitted.

The outcomes of the RAE are published and provide public information about the quality of research in UK universities and colleges. It is useful for the voluntary sector, industry and commerce to guide their research funding decisions. The RAE also gives an indication of the relative quality and standing of UK academic research and provides benchmarks that are used by institutions in developing and managing their research strategies. Research quality as measured by the RAE has improved dramatically over the last decade and it is argued that the RAE tends to lead to concentration of research and a proliferation of research papers. There are calls from within the academic community for future RAEs to be metrics-based (Harnard, 2006)

For this thesis, the average RAE rating of the UK universities was taken from the Times Higher Education Supplement (Mayfield University Consultants, 2001), which averages the grades awarded to each university by the Government RAE. Other research has already used this data effectively, revealing statistically significant correlations with hyperlinks (Smith & Thelwall, 2002; Thelwall, 2002a; Thelwall, 2003b; Li et al., 2003).

4.7 Academic Staff Numbers

Since the subject matter of the early part of the current research has important implications for the design of web site metrics, and also as it is the measure used in most of the linear trend-based papers employing UK link data, it seems natural to use a definition of a research activity indicator as academic staff numbers multiplied by the average RAE score for each individual university. On occasion, this is used in preference to web site page counts, as site page counts have proved to provide unreliable data (Thelwall, 2001d).

Staff numbers were taken from the Noble Publishing Co. (1999) and, while it is recognised that the RAE ratings were derived from 2001 data and that the combination of data from different years is not ideal, staff numbers and RAE averages are relatively stable for most universities and so should not significantly impact upon the results.

4.8 Data Validation

Irrespective of any data collection technique used, it is critical that the data is assessed for its validity. In other words, does the research truly measure that which it was intended to measure? Data validation involves a process to determine if the data is accurate, complete, or meets specified criteria. An analogy may be made between this and the validation of data to be used in computer programs and databases. In both cases, data validation ensures that the data is sensible before it is processed, either automatically or manually.

Researchers generally determine validity by asking a series of questions, and will often look for the answers in the research of others. In this case, the raw data used in this study are the hyperlink text files of UK, Australian and New Zealand universities. This data has already been used successfully in many previous web-based research (chapter 2.3.4.2), and this supports the case for its validity.

Correlation testing is used extensively during this thesis. Correlation summarises the strength of the relationship between two variables and, while several different correlation coefficients can be calculated, the two most commonly used are Pearson’s correlation coefficient and Spearman’s rank correlation coefficient. Pearson’s correlation coefficient requires both variables to be measured on an interval or ratio scale and the calculation is based on the actual values. Spearman’s rank coefficient requires data that are at least ordinal and the calculation, which is the same as for Pearson’s correlation, is carried out on the ranks of the data (each variable is ranked separately by putting the values of the variable in order and numbering them).

To perform correlation tests using Pearson’s correlation coefficient it is necessary to assume that both variables have a normal distribution but no such assumption is necessary for tests on Spearman’s rank correlation. For normal distributions, Pearson’s correlation coefficient has a slight advantage over Spearman’s rank correlation coefficient and provides the most powerful significance test of correlation. However, several previous web based studies (Tang & Thelwall, 2003; Thelwall & Tang, 2003; Thelwall & Price, 2003; Thelwall & Harries, 2004a; Thelwall, 2004b) have subjected the raw data used during this study to the Kolmogorov–Smirnov test, (the principal goodness of fit test for normal and uniform data sets), and the web data used was found to be significantly non-normal. Therefore Spearman’s coefficient is preferred to Pearson’s coefficient throughout the course of this thesis. This is an important methodological issue as non-parametric Spearman’s correlation tests are normally more appropriate than Pearson’s, given the typically skewed nature of web link data. (Pearson’s correlation coefficient is parametric while Spearman’s is non-parametric. Parametric tests are those designed for normally distributed data while non-parametric tests are designed for data that is not normally distributed; the Chi-Square test of independence, used extensively in chapter 7, is another example of a non-parametric test).

Webometric research must be conducted with caution as the data collection instrument (the SocSciBot crawler), as well as the data source (the hyperlink structure of the UK, Australian and New Zealand academic webs), may have deficiencies. Therefore, another important aspect of assessing the validity of the data is to assess the validity of the data collection method. Issues directed at the appropriateness of the data collection method, its reliability and the representitiveness of the sample should be identified and addressed. In this case, the main data collection method is the SocSciBot crawler and, while this has advantages over the use of commercial search engines when collecting university web data, it does present its own, specific issues.

The fact that a complete data set of all UK, Australian and New Zealand university links is available precluded the use of random samples, as used in other longitudinal studies. However, although attempts were made, especially from 2001 onwards, to crawl each national academic web space every year, and at roughly the same time each year, these crawls were not taken at exactly the same time period each year. Having said that, given the similarities between each academic web space, the slight time discrepancy seems unlikely to significantly affect the results.

Another concern was the fact that the data collection method, by design, only crawled static web pages. Therefore, effort was made to identify and display the trends associated with the three academic webs use of dynamically generated web pages and non-HTML documents. Additionally, while effort has been made to validate the findings by producing results accounting for embedded links and dynamically generated web pages, there may be a deficiency in the data collection instrument in its inability to count obscured links. As discussed in the literature review, obscured links are URLs that can be accessed by web users in ways that are difficult or impossible for web crawlers. The link extractor part of a crawler is not capable of extracting all links from web pages because some can be stored in formats that are in practice impossible for them to decode. For example, with programs running through the web browser, such as JavaScript, Java, Shockwave and Flash, it is not possible to easily extract URLs since these may be built by the code itself when running. This would mean a site using this kind of technology without the back up of HTML links would not be covered completely.

As well as addressing concerns with the validity of the raw data, this thesis also addresses the validity of the results. In each area of research, a systematic review of the results was carried out to identify outliers or suspect values. This process was intended to screen out highly unlikely values which had the potential to skew the results. The Internet Archive, despite an unintentional international bias in its coverage, is an important tool, especially for a longitudinal webometric study and was used extensively to clarify and consolidate any ambiguous results.

Many webometric studies use inlink counts in preference to outlink counts as it is often argued that inlinks are more useful as indicators than outlinks, as outlinks are under the control of the site owners whereas inlinks are not. An additional technical problem with site outlink counts is that they depend upon a single site crawl, and are therefore more liable to crawler coverage problems than inlink counts, which are totalled from a number of different crawls. In order to avoid the problems associated with the used of either inlink or outlink counts, this study uses both.

The validity of the non-web data used is also not beyond question. The RAE rating figures used are now over five years old, and the methods of counting academic staff numbers may vary considerably between universities. The staff numbers, RAE ratings and web link data were derived from different years and, while it is recognised that this is not ideal, especially for a longitudinal study, staff numbers and RAE averages are relatively stable for most universities and so should not significantly impact upon the results.

The above discussion shows a fundamental methodological problem in webometrics. It must handle data of a much more messy, non-standardised, diverse and dynamic nature than traditional bibliographic data used in bibliometrics and scientometrics, even though data validation also is required in these fields in order to obtain adequate comparable units of data as a basis for empirical investigations.

4.8.1 Data Reliability

Another important consideration with regard to any chosen data set is its reliability. Data reliability is defined as the extent to which results are consistent over time and are an accurate representation of the total population under study. In other words, if the results of the study can be reproduced under a similar methodology, then the research instrument is considered to be reliable.

This study, by its very nature, examines changes over time. While the point of the exercise is to identify and explain any changes, the changes themselves, although showing significant and sometimes surprising results, are not impossible or even highly unlikely, year-on-year. The quantitative, empirical nature of many of these results, using computer programs to identify mathematical patterns and trends within hyperlink data, naturally lends itself to a high degree of reliability.

In addition, in the early days of this research, many of the results were produced using bespoke, special-to-type Visual Basic programs. These results were later compared, and confirmed, using the SocSciBot suite of tools. Using different tools to obtain similar results helps to confirm the reliability of both the methodology and the raw data.

Finally, although the current study seeks to make a significant new contribution to the existing knowledge base, some of the research techniques used involve the extension of existing research to other areas. In particular, at times the research will take results from a single academic web and apply it to others, or take established findings and examine how they change over time. During this process, where the original results are often confirmed, this element of repeatability again confirms the data's reliability.

4.9 Summary

This chapter details the varied, numerous data collection methods used during the course of this research, and attempts to highlight some of the advantages and disadvantages of each. These data collection methods were then assessed in an attempt to evaluate of the validity and reliability of the data and results, showing that they appear to have a high level of validity and a very high level of reliability.

5 Academic Web Models: Linear Relationship or Non-Linear Power Law?

5.1 Introduction

Several previous studies have used the WIF to investigate the relationship between interlinking counts of a selected set of universities on a national level and the research of those universities. Positive correlations were found between research measures and inlink counts for universities in the UK (Thelwall, 2001d, 2002e), Australia (Smith & Thelwall, 2002), Taiwan and Mainland China (Thelwall & Tang, 2003) and Canada (Vaughan & Thelwall, 2005). Most research into academic web interlinking using the WIF has tended to assume that web data should follow a linear trend (Thelwall, 2002e; Thelwall, 2002d; Thelwall & Wilkinson, 2003b). However, if academic web data is shown to demonstrate power law behaviour, then it will affect the design of web metrics such as the WIF, and calculations based on web metrics would have to be modified to take this into account.

Significant relationships and patterns can be extracted from the analysis of web links between academic institutions and it is now becoming accepted that it is important for researchers, and the academic community in general, to benefit from mining the linking of scholarly materials on the web in various ways (e.g. Jepsen et al., 2004). Nevertheless, the question of how best to extract patterns from web links remains problematic in some respects. Previous studies of academic web interlinking have tended to hypothesise that the relationship between the research of a university and links to its web site should follow a linear trend, yet the typical distribution of web data, in general, seems to be a non-linear power law (see chapter 2.4.1).

The aim of this chapter is to assess whether a linear trend or a power law is the most appropriate method with which to model the relationship between academic web size, interlinking data and research activity indicators. This was considered to be an important question, and one which should be determined at the very outset of the current study as, if the academic web were shown to demonstrate power law behaviour, it would have implications for previous and future web metric based research because specially modified 'scale-independent' indicators would then need to be developed (Katz, 2000).

5.2 Research Question

Previous research into academic web interlinking has tended to hypothesise that the relationship between links to a university and its research should follow a linear trend (Thelwall, 2002e; Thelwall, 2002d; Thelwall & Wilkinson, 2003b). Yet the distribution of links to individual academic pages follows a power law (Thelwall & Wilkinson 2003c). What has not been conclusively demonstrated however, is that the relationship between link counts, when aggregated at the university level, and research indicators follows a linear rather than a non-linear power law. Put another way, should the links that a university creates or attracts always be proportional to a measure of its research, or should larger universities expect to attract or create more (or less) links than this, by virtue of their size?

If the link count-research relationship does follow a power law, then this has important implications for the design of web metrics such as the WIF (Ingwersen, 1998). For example, Katz (2000) has shown that specially modified ‘scale-independent’ indicators need to be developed for data that obeys a power law (see also Leydesdorff & Bensman, 2005). If power law behaviour were to be exhibited, then calculations based on WIFs would have to be modified to take this into account.

The primary research question is therefore as follows:

Is the relationship between academic web and research activity indicators best modelled by a linear trend or a power law?

In order to answer this question, this chapter will concentrate on the following more specific sub-question:

For UK university web sites, is the relationship between site size and research indicators, and between outlinks and research indicators, best modelled by a linear trend or a non-linear power law?

5.3 Methods

5.3.1 Raw Data

The raw link data used during this study was derived from the text files of 111 UK universities as of June-July 2002, the link structure of which was obtained by the specialist information science web crawler SocSciBot and made available in Database 9 of the Wolverhampton University Academic Web Link Database Project (Thelwall, 2002/3).

The link structure of Database 9 consists of a separate text file for each UK university, giving a list of the URLs of all source pages crawled together with all identified target URLs referred to in the page, with duplicate URLs removed and all URLs truncated at the first ‘#’ character (Thelwall, 2002e).

The university text files were processed using a specially written program which sorted the link data, removed all duplicates and aggregated the resultant data, i.e. source URL counts and target URL (link) counts from each university crawl, using ADMs. The basic data sets for each site can be found at Appendix 1. Note that site self-links are included in the data set.

5.3.2 Alternative Document Models

The ADMs used within this chapter aggregate pages at web page, directory and domain levels using the standard ADM definitions given in chapter 2.3.4.1.2.

5.3.3 Staff Numbers and RAE Ratings

Fundamental to the research question is the definition of ‘academic web size’. Previous studies have used a number of factors in order to describe university site size, including the number of web pages within the university source domain, the number of university staff and the physical size of the university.

Following the work of Ingwersen (1998) in creating the external relative WIF, a version of this external relative WIF was created especially for academic web sites (Thelwall, 2001f). This uses the number of full-time academic members of staff at the university instead of a site page count as the denominator of the calculation.

WIF metrics using this calculation, together with UK RAE derived data have been previously utilised revealing statistically significant correlations (Smith & Thelwall, 2002; Thelwall, 2002a; Thelwall, 2003b; Li et al., 2003). Since the subject of the research question in the current study has important implications for the design of web site metrics, and also as it is the measure used in most of the linear trend-based papers employing UK link data, it seems natural to use the definition of research activity as academic staff numbers multiplied by the average RAE score for each individual university. This is used in preference to web site page counts, as site page counts have proved to provide unreliable data (Thelwall, 2001d).

The average RAE rating of the universities was taken from the Times Higher Education Supplement (Mayfield University Consultants, 2001), which averages the grades awarded to each university by the government RAE. This is a peer review, subject-based process that is used to direct Government research funding. Staff numbers were taken from the Noble Publishing Co. (1999). The combination of data from different years is not ideal, although staff numbers and RAE averages are relatively stable for most universities and so should not significantly impact upon the results. University full-time academic staff numbers and RAE ratings can be found at Appendix 2.

5.3.4 Statistical Analysis

The statistical analysis program SPSS 10.0 was used to perform a number of tests on both the raw and ADM aggregated data, including linear regression and correlation analysis.

Linear regression analyses the relationship between two variables, x and y. For each subject, both x and y are known, and it is required to find the best straight line through the data. In general, the aim of linear regression is to adjust the values of slope and intercept to find the line that best predicts y from x. More precisely, the goal of the regression is to minimise the sum of the squares of the vertical distances of the points from the line. Note that linear regression does not test whether the data is linear. It assumes that the data is linear, and finds the slope and intercept that make a straight line best fit the data.

The equation of the power type trend line displayed in logarithmic graphs takes the form $y = cx^b$. That is, y is a power law in x with a power or index of b multiplied by a number or normalisation constant c.

Now, if it can be shown that the linear power 1 is within the 95% confidence interval for the power law parameter for the equation index, this can be used to claim that there is no evidence of a power law. This is because if $y = cx^b$, where $b = 1$, then $y = cx$, which is a linear equation, and this would mean that the relationship between UK university link data and research activity indicators would be best modelled by a linear trend, and not a power law.

It is worth noting that there are direct techniques for testing for power laws, and there is software available to easily do this (Rousseau & Rousseau, 2000). However, the current paper does not directly test for a power law, as a power law fit could also be a linear trend fit and so would not answer the primary research question.

5.4 Results

The 111 raw data university text files were analysed and aggregated into directories and domains using ADMs, and statistics produced for source and target domains, directories and pages for each individual university. These were taken as dependent variables and plotted on six scatter graphs against the 'staff number * RAE rating' data as independent variables. The data was previously subjected to the Kolmogorov-Smirnov test, the principal goodness of fit test for normal and uniform data sets, and the test distribution was found to be not significantly non-normal.

Figures 5.1 to 5.6 show the data on a linear scale and Figures 5.7 to 5.12 show the same data on a logarithmic scale.

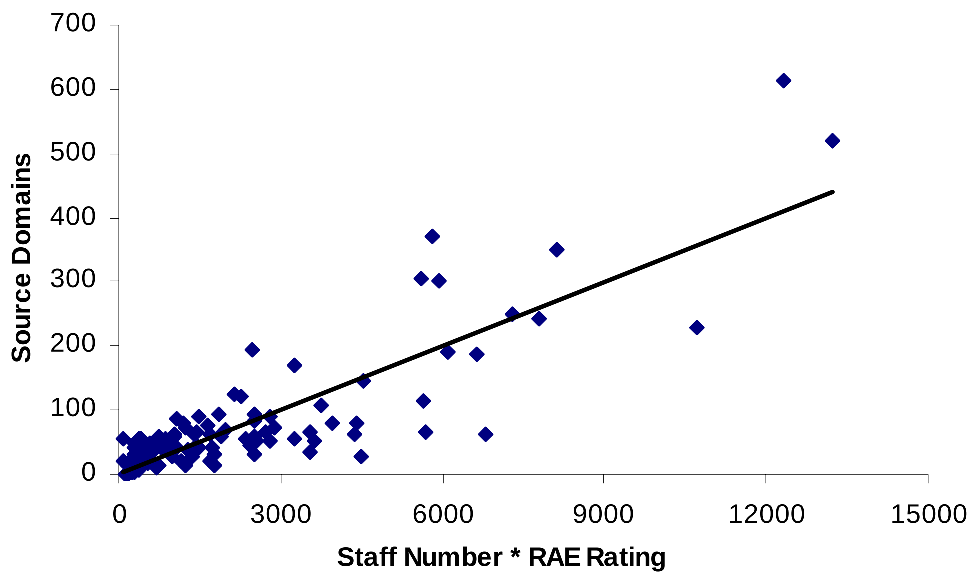

*Figure 5.1 Source Size (number of Domain ADMs in each site) against Research (Staff Number * RAE Rating)*

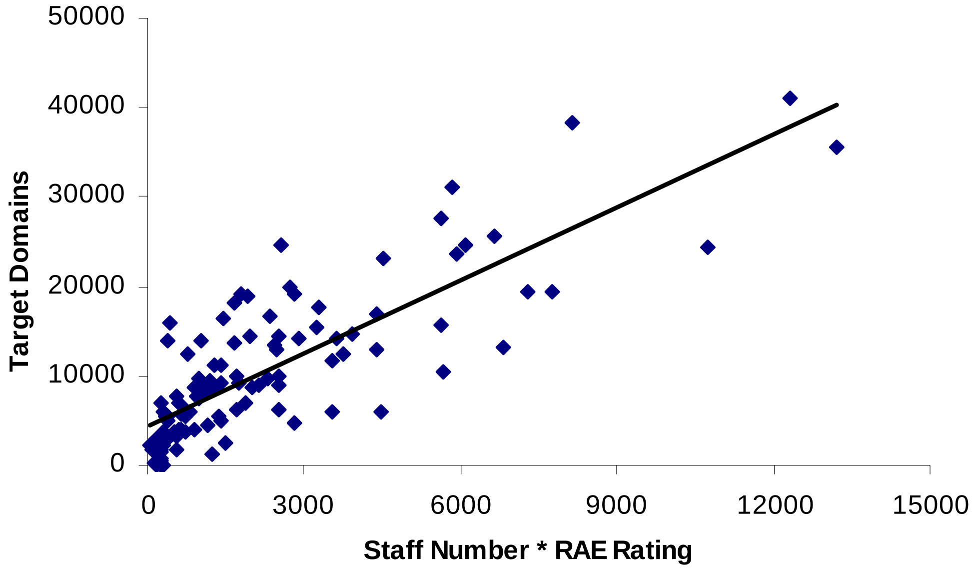

*Figure 5.2 Target Size (number of Domain ADMs targeted by links from each site) against Research (Staff Number * RAE Rating)*

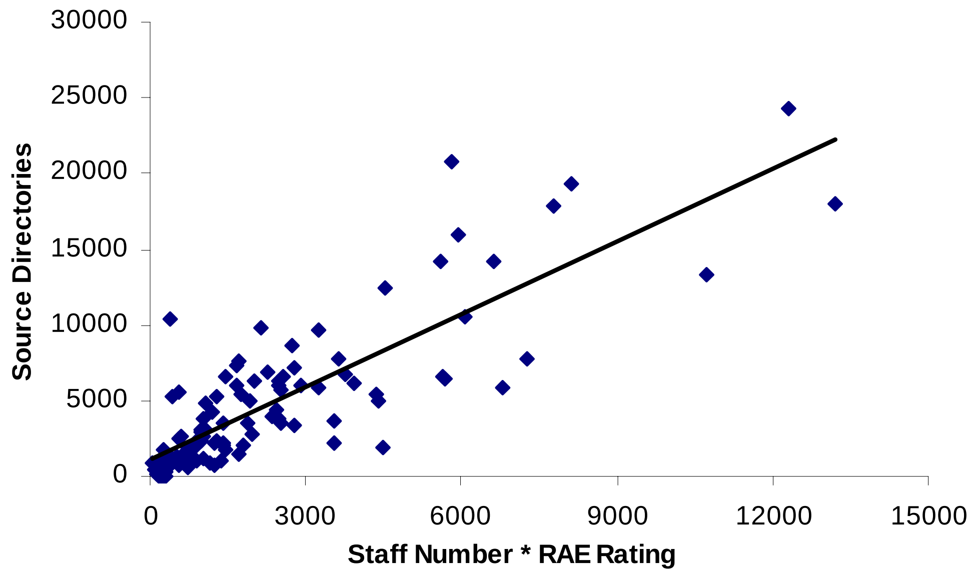

*Figure 5.3 Source Size (number of Directory ADMs in each site) against Research (Staff Number * RAE Rating)*

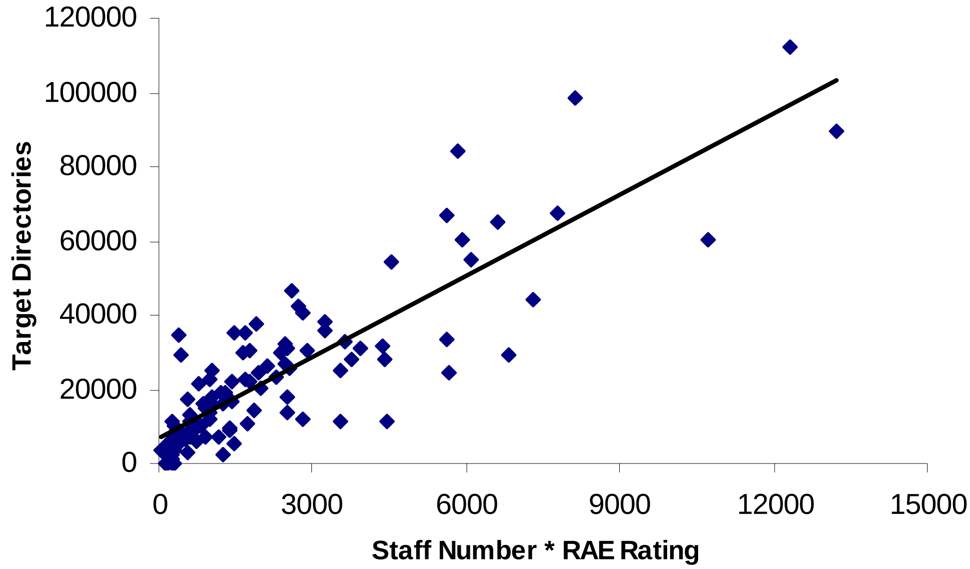

*Figure 5.4 Target Size (number of Directory ADMs targeted by links from each site) against Research (Staff Number * RAE Rating)*

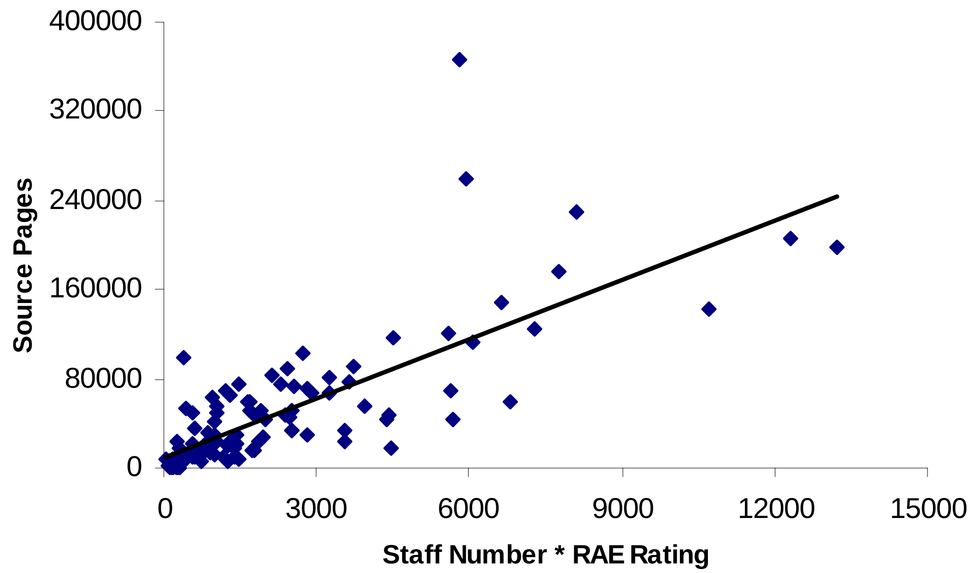

*Figure 5.5 Source Size (number of Page ADMs in each site) against Research (Staff Number * RAE Rating)*

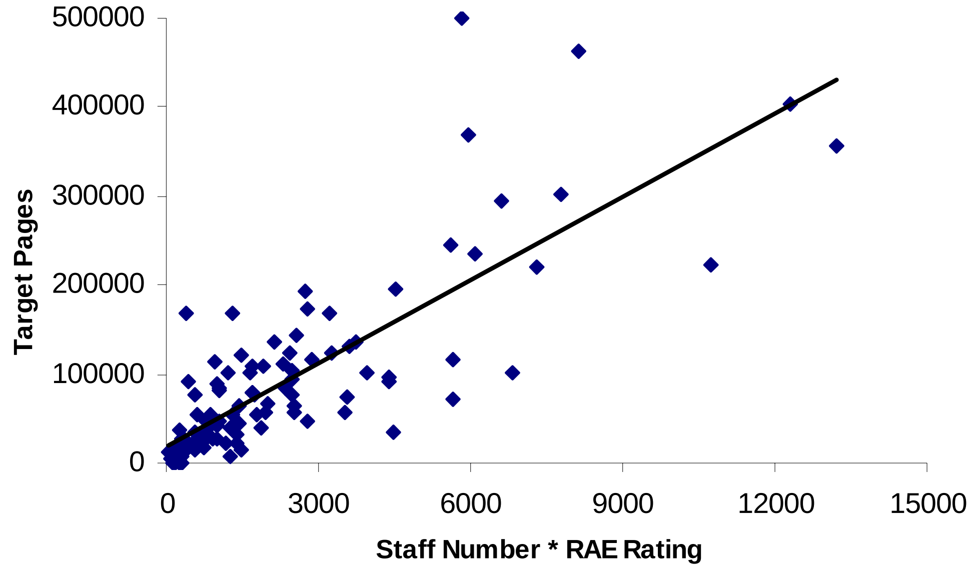

*Figure 5.6 Target Size (number of Page ADMs targeted by links from each site) against Research (Staff Number * RAE Rating)*

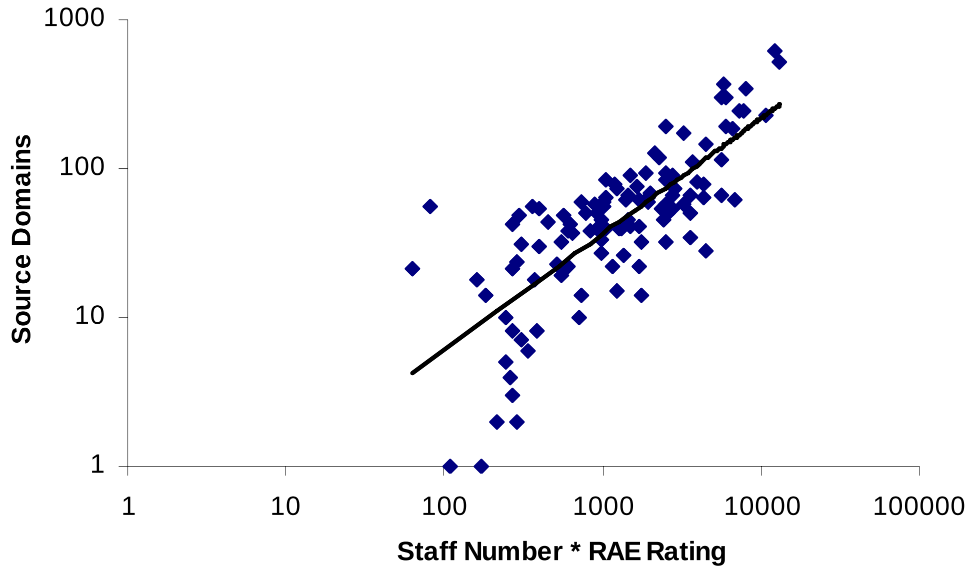

*Figure 5.7 Logarithmic graph of Source Size (number of Domain ADMs in each site) against Research (Staff Number * RAE Rating)*

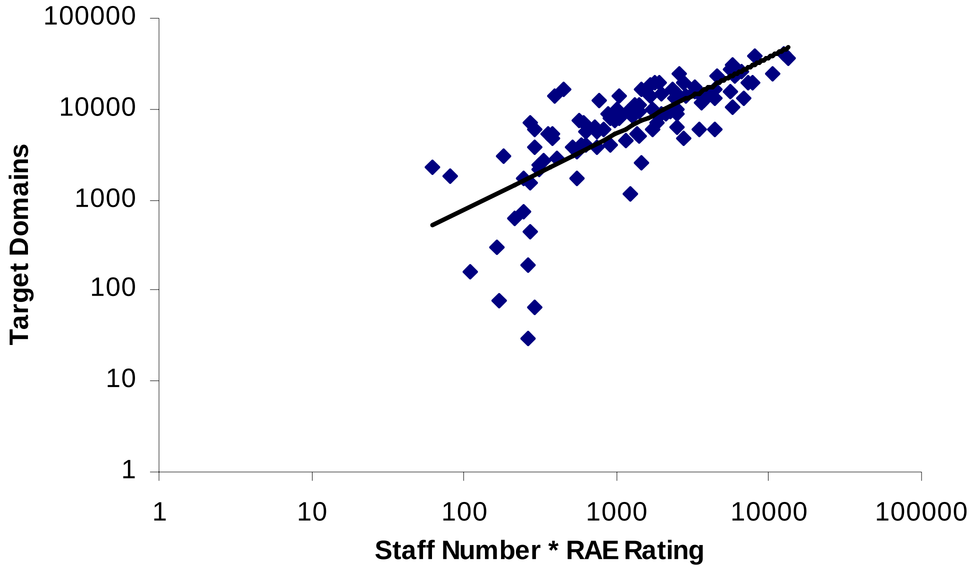

*Figure 5.8 Logarithmic graph of Target Size (number of Domain ADMs targeted by links from each site) against Research (Staff Number * RAE Rating)*

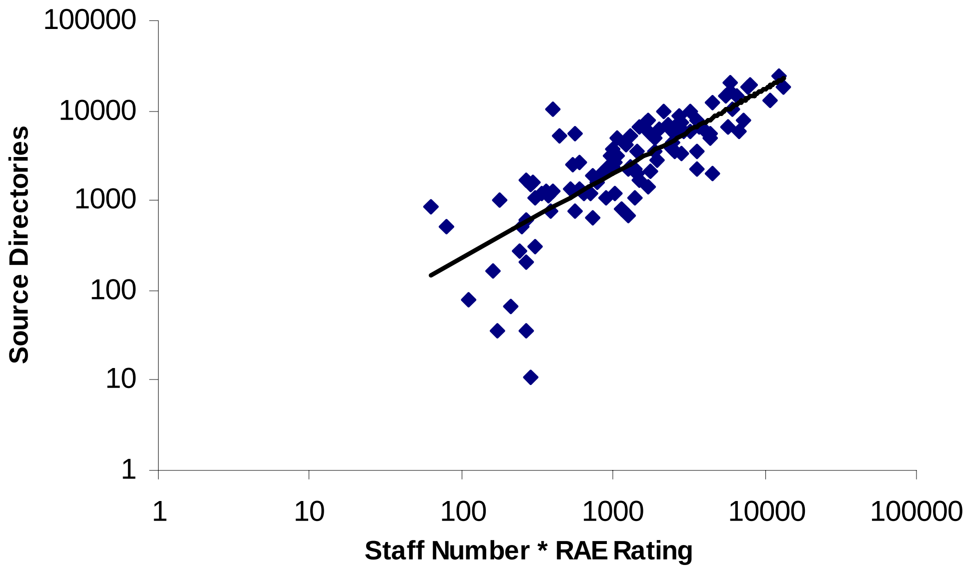

*Figure 5.9 Logarithmic graph of Source Size (number of Directory ADMs in each site) against Research (Staff Number * RAE Rating)*

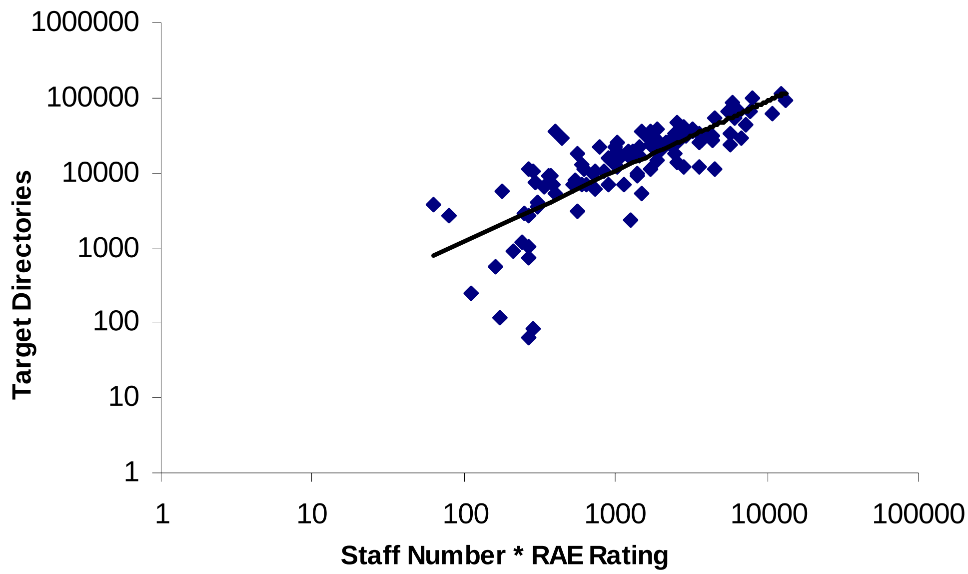

*Figure 5.10 Logarithmic graph of Target Size (number of Directory ADMs targeted by links from each site) against Research (Staff Number * RAE Rating)*

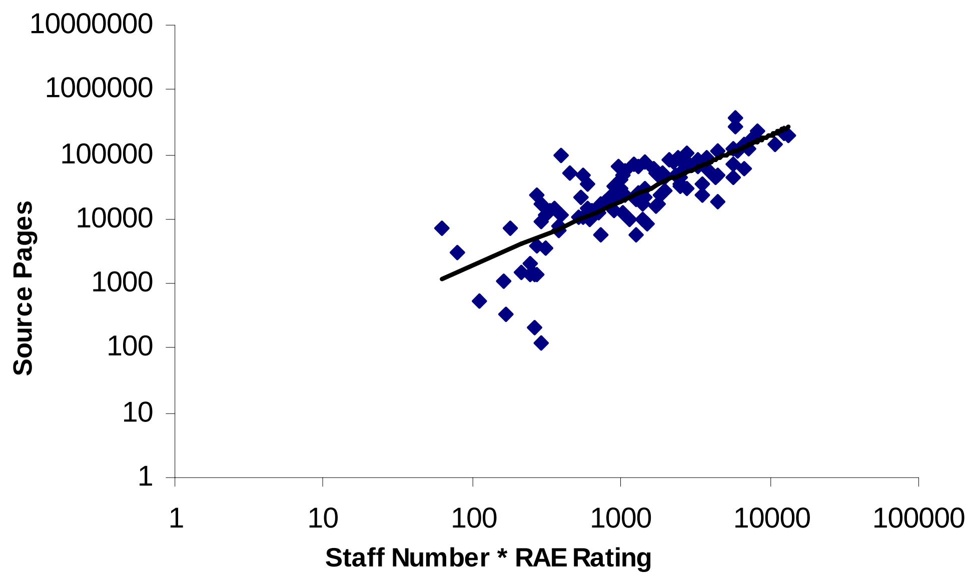

*Figure 5.11 Logarithmic graph of Source Size (number of Page ADMs in each site) against Research (Staff Number * RAE Rating)*

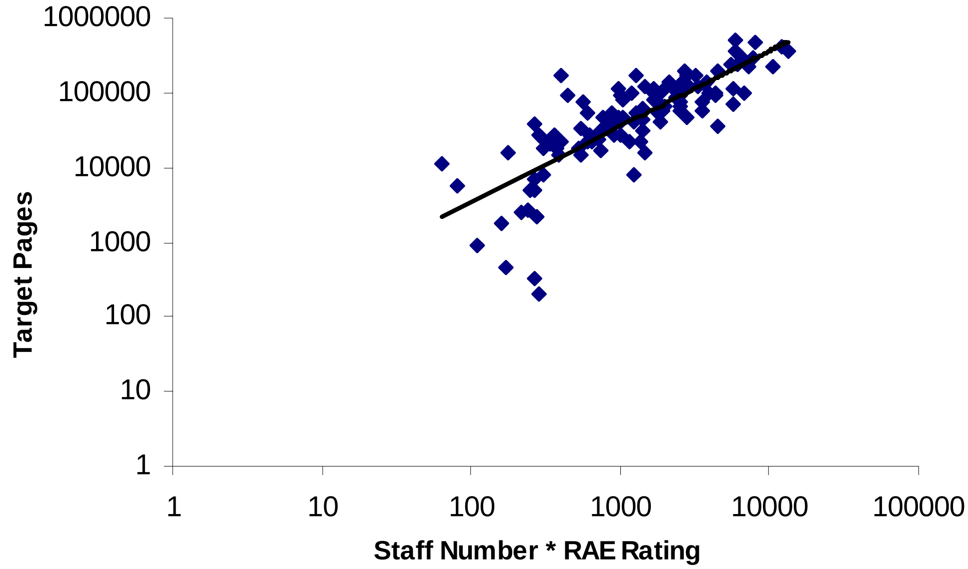

*Figure 5.12 Logarithmic graph of Target Size (number of Page ADMs targeted by links from each site) against Research (Staff Number * RAE Rating)*

In carrying out linear regression from a power law, let (x, y) be the raw data and suppose that a linear regression equation is fitted to the logged data (ln x, ln y). Then ln $y = a + b$ ln x, which can be written as $y = e^a x^b$. We can see that a non-zero value of a does not imply that y is not linear in x, only that y is proportional to x^b rather than equal to x^b.

However, there is an implicit assumption that the straight line fit by the original linear data passes through zero (0, 0) as the logarithmic data must pass through zero. If the underlying data does not pass through zero then test is invalid because the power law line $y = e^a x^b$ could never fit it.

Table 5.1 y-axis intercepts for the line slopes in Figures 5.1 – 5.6

ADM	Source documents	Target documents
Domain	2.7952 (p = 0.668)	4369.8 (p = 0.000)
Directory	1037.6 (p = 0.002)	6477.6 (p = 0.000)
Page	9718.8 (p = 0.046)	18291 (p = 0.014)

Table 5.1 above gives the results of applying linear regression to the original data. The fact that all graphs with the exception of Figure 5.1 give the equivalent p value as being less than 0.05 should not be ignored and may, to a certain extent, undermine the validity of the tests. However, in context with the raw data, the linear fit does not have an intercept significantly different from zero.

For the cases where the intercept of the origin for the linear regression line is significantly non-zero (i.e., all except one of the cases), as the intercept is non-zero, it can be observed that no power law can fit the data (as all power laws have pass through the origin in unlogged form). This can be seen as a useful double-check for the existence of a power law.

Visual inspection of graphs 5.1 – 5.12 suggests that linear trends and power laws both fit the data sets to a reasonable extent, although it is also clear that the data does not exhibit the very marked power law pattern seen in previous web exercises, in which the power law slope is in the opposite direction to the above graphs and shows a marked 'broomstick' or 'hooked broomstick' shape, reflecting the 'rich get richer' phenomena. The graphs above resemble more those of Katz (2000), pointing to the possibility that there are economies of scale operating. However, to prove whether or not the best fit for these graphs is a linear or logarithmic line, more rigorous tests need to be applied.

95% confidence intervals were taken for the power in the power laws for each data set. This means that, given the assumptions of linear regression, the confidence interval is 95% sure to contain the best-fit regression line, leaving a 5% chance that the true line is outside those boundaries. However, this is not the same as saying that it will contain 95% of the data points. The results for each of the logarithmic graphs are shown in the table below.

Table 5.2 Upper and Lower 95% Confidence Intervals for Power Law Powers (the line slopes in Figures 5.7 – 5.12)

ADM	Source documents	Target documents
Domain	(0.652, 0.901)	(0.698, 0.986)
Directory	(0.798, 1.077)	(0.786, 1.074)
Page	(0.863, 1.162)	(0.863, 1.157)

Table 5.3 gives Spearman's correlation coefficient, ρ, which gives a measure of linear association.

Table 5.3 Spearman's Correlation Coefficient for Figures 5.7 – 5.12

ADM	Source documents	Target documents
Domain	0.776	0.816
Directory	0.834	0.842
Page	0.807	0.829

All graphs display correlation coefficients in a very narrow range between 0.776 and 0.842. Levels of correlation between 0.7 to 0.9 can be described as high and so, using Spearman's nonparametric correlation coefficient, ρ, it is clear that the graphs displayed in Figures 5.7 – 5.12 all display high levels of correlation, all significant at the 0.01 level (2-tailed).

5.5 Discussion

5.5.1 Power Law or Linear Trend?

It can be observed from Table 5.1 that the linear power 1 does lie within the 95% confidence interval for the graphs of Source Directories vs. Staff Number*RAE Rating, Target Directories vs. Staff Number*RAE Rating, Source Pages vs. Staff Number*RAE Rating and Target Pages vs. Staff Number*RAE Rating. So it could be said that, for these graphs, it would be reasonable to use a linear model for the relationship. However, the linear power 1 does not lie within the 95% confidence interval for the graphs of Staff*RAE vs. Source Domains and Staff*RAE vs. Target Domains, although upper 95% confidence intervals of 0.901 and 0.986 respectively suggest that there may be evidence of linear trends which may be masked by inaccuracies in the data, inadvertently introduced as a result of the limitations highlighted in chapter 5.5.3.

The high levels of correlation shown in the results, significant at the 0.01 level (2-tailed), do seem to suggest a definite linear relationship between the Staff Number*RAE Rating data and the number of source domains, target domains, source directories, target directories, source pages and target pages, and go some way to justify the choice of methodology used.

It should be noted that all six logarithmic graphs show the same basic pattern with the two graphs utilising the domain ADMs containing the most extreme outliers. These are also the graphs that show the most deviation from the 95% confidence intervals and the lowest levels of correlation.

5.5.2 Outliers

The following universities were noted as reoccurring outliers for the corresponding charts:

Paisley. This is not only the smallest university in terms of the number of subdomains but also has the smallest number of source directories, source and target pages.
Thames Valley University (TVU). TVU has the lowest average RAE rating of 0.4, and this contributes to the lowest Staff Number*RAE Rating.
Newport. This university has the joint second lowest RAE rating (0.5), and the second lowest Staff Number*RAE Rating.
Chichester. Although not the smallest university, it does have the smallest number of both target domains and directories. The large number of source page links observed can be explained by the design of its web page menu structure.

Because of the combination of low numbers of full-time academic staff and low average RAE scores, TVU and Newport have the lowest Staff Number*RAE Rating (63.04 and 80.15 respectively) by some margin – the next lowest being Anglia with 361.1. The removal of the outliers with the lowest RAE ratings, i.e. Newport and TVU universities, brings the 95% confidence interval for the Target Domains vs. Staff Number*RAE Rating graph to 1.048 and the 99% confidence interval for the

Source Domains vs. Staff Number*RAE Rating graph to 1.026, i.e. the linear power 1 is now within the respective confidence intervals for both of these graphs.

The removal of outliers is considered to be a permissible step in statistical analysis, and in this case, does not change the impact of the conclusions.

5.5.3 Limitations

There are a number of limitations of this study, some of which are highlighted in the list below:

- Universities have different policies for domain name use (Thelwall & Harries, 2003). This may explain the deviation in the source and target domain 95% confidence intervals.
- The study only covers one national university system. It is possible that domain structure in other countries would be different, and so the results would not necessarily extend. Future results confirming the results for other countries would strengthen the findings.
- The study covers only web site size and outlink counts. Although it seems unlikely, it is possible that different results could be obtained from site inlink counts, which are most commonly used in webometrics, as well as site outlink counts (i.e. excluding site self-links), and the interlinking counts used in the WCF calculation.
- The validity of the non-web data used is not beyond question. The RAE rating figures used are now a number of years old, and the methods of counting academic staff numbers may vary considerably between universities.
- The staff numbers, RAE ratings and web link data were derived from three different years (1999, 2001 and 2002 respectively). While no significant change in the data is expected, minor inaccuracies may be introduced as a result of this time-span.
- There may be inaccuracies within the hyperlink structure web data such as typographical errors in target URLs.

These are clearly drawbacks that should encourage caution in the interpretation of the conclusions as these limitations, some of which would be inherent to any similar web site counting study, make it much harder to identify mathematical patterns, even when they are actually present. It is hoped that future web link research will improve the reliability of the data and increase confidence in the results but until then, the evidence presented here is supportive of the hypothesis that the relationship between academic web size and research activity indicators is best modelled by linear trends.

5.6 Conclusions

This chapter has taken as raw data the text files of UK university hyperlink structures, aggregated them into source and target domain, directory and page ADMs, and displayed the results against academic staff numbers multiplied by average RAE rating data on both linear and logarithmic graphs in an attempt to show whether a linear trend or a power law is the most appropriate method with which to model the relationship between web publishing/linking and research activity.

Following linear regression, analysis of the logarithmic graphs showed that the linear power 1 fell within the 95% confidence interval for the following graphs:

- Number of Source Page ADMs vs. Staff Number*RAE Rating
- Number of Target Page ADMs vs. Staff Number*RAE Rating
- Number of Source Directory ADMs vs. Staff Number*RAE Rating
- Number of Target Directory ADMs vs. Staff Number*RAE Rating

It is a reasonable interpretation then, that this data, using the page and directory ADM, is broadly consistent with a linear trend. The results using the domain ADM were not as conclusive, although even with no further analysis, with upper 95% confidence intervals of 0.901 and 0.986 for source domain ADM and target domain ADM respectively, it could be claimed that a linear trend is still in evidence. The deviation could be explained by inaccuracies in the raw data collection method, application of the ADM models, use of the staff number / RAE data or a combination of all three.

The situation for the domain ADM is different. The linear power 1 fell outside of the 95% confidence interval and there was some evidence of a non-linear power law, even after the removal of prominent outliers. Inspection of the graph, however, does not show a clear power law trend; the clearest non-linear factor is that low research activity universities have few domains. A non-linear power law would explain this, but an alternative explanation would be a break-point: perhaps universities need a critical research size in order to create the infrastructure (technical knowledge, and authority permission) to be able to create new subdomains as and when needed. In contrast to domains, directories typically do not need specialist knowledge, permission and access to create and so are in a sense more natural. This argument supports the use of the directory ADM as preferential to the domain ADM, at least until this apparent anomaly in domain creation is explained or disappears.

The fact that the relationship between academic web site size and research activity is best modelled by a linear trend validates the results of previous research into academic web interlinking which has assumed that web data should follow a linear trend (Thelwall, 2002e; Thelwall, 2002d; Thelwall & Wilkinson, 2003b). It also justifies the design and use of web metrics such as the WIF (Ingwersen, 1998), which rely on linear data. Katz (2000) has shown that specially modified 'scale-independent' indicators need to be developed for data that obeys a power law. However, many of the ‘rich gets richer’ type of power law behaviour observed within the web concentrate on the distribution of overall web links over time, and is not strictly applicable to the collection of academic text file hyperlink structures used within this research, as they represent a snapshot of university hyperlink data.

The results from this study hold importance as a step towards understanding the phenomenon of academic web linking and developing metrics to extract useful information. A greater understanding of the mathematical patterns and relationships within the hyperlink structure of these links will develop an appreciation of the way the academic web is presently connected, and may prove to be a useful tool in predicting future development and evolution.

It is evident that academic hyperlink analysis can be complex and problematical. Although significant mathematical patterns can be extracted from hyperlinks, it is still the case that they are a largely unregulated phenomenon. As a result great care must be taken to validate data when conducting hyperlink analyses to avoid drawing false conclusions because of data unreliability. Nevertheless, these positive results strengthen the case for using web link analysis as a tool with the potential to reveal underlying trends in academic web site interlinking.

Despite the problems mentioned above, the confidence intervals for the respective ADMs vs. Staff Number*RAE Rating data are statistically significant. The positive results can confirm that it is a linear trend, and not a power law, which best models academic web size and interlinking data.

6 A Longitudinal Analysis of Alternative Document Models

6.1 Introduction

ADMs have been used extensively in academic web research as aggregated units of analysis. They were compared with estimated research activity indicators for UK university institutions and produced significant results, giving additional support for a closer relationship between links and research for the domain and directory ADMs, with the latter being the method of choice (Thelwall, 2002e). Very strong correlations were also observed by restricting inlink counts to target pages mainly connected with research (Thelwall and Harries, 2003), and this gave a high degree of confidence that links between university web sites are connected with scholarly activity in some way, despite the number that are created for recreational reasons (Wilkinson et al., 2003). Thelwall and Aguillo (2003) utilised ADMs to undertake a health check of Spanish universities while Thelwall et al., (2003) calculated the percentage of the highest inlinked subject-based web sites of universities in Taiwan and Australia in 2003, classifying the sites found by crawling the university web sites in each country, applying the domain ADM to their link structures and then selecting the 100 highest inlinked subject-based web sites in each country. ADMs were also used to show significant correlation between university inlinks and research in Thailand, but the results were inconclusive for Mainland China, possibly due to scoring inadequacies (Thelwall and Tang, 2003), and to show that faculty quality and language are important predictors to links to Canadian university web sites (Vaughan and Thelwall, 2005).

Payne and Thelwall (2004) presented the results of statistical analysis carried out on the web link structure text files of 111 UK universities during 2003. Summary statistics were produced using ADMs and the results of the statistical analysis were also graphically displayed, using both linear and logarithmic graphs, including trendline equations. Mathematical linear relationships were observed between certain bivariate data with subsequent correlation analyses revealing a number of very strong relationships, particularly between site size and number of source / target directories and pages. However, a noted weakness in this methodology was that the results presented only concerned one national university system, crawled at one time and was clearly a drawback that should encourage caution in the interpretation of the conclusions in other contexts.

This chapter seeks to expand on this study by means of increasing its scope to include the New Zealand and Australian academic webs and to introduce a longitudinal aspect, looking at which ADM proves to be the most reliable over a six year period.

6.2 Research Question

The results so far show that meaningful information can be extracted from large scale comparisons of academic web links and, against this background of research, it seems natural to investigate whether or not the validity and accuracy of ADMs change over time and whether their proven ability to reduce the effect of anomalies for UK university data extends to other academic web spaces.

More specifically, this paper, which is the first longitudinal study involving ADMs, addresses the following research question:

Which Alternative Document Models give the most consistent results when applied to the UK, Australian and New Zealand academic web spaces 2000 – 2006?

In this context, consistency is measured by assessing which ADMs, when examined over time, reliably show the greatest levels of correlation with a standard measure of university site size. The underlying assumption is that the number of links to a site, if appropriately measured, should be proportional to the size of the site.

6.3 Methods

A specialist webometrics crawler (Thelwall, 2001a) initially collected the raw data and stored the hyperlink structure text files of each of the UK, Australian and New Zealand universities in publicly available databases as part of the University of Wolverhampton Academic Web Link Database Project (Thelwall, 2002/3). Several papers have already been published based on results produced using this information.

Freely available information science software (SocSciBot Tools, socscibot.wlv.ac.uk) was then used to process these text files and produce standard page, directory, domain and site (or university) ADMs using both inlinks and outlinks. The four standard ADMs used in this chapter are defined previously in chapter 2.3.4.1.2 and the raw data used in this chapter can be found in Appendices 3 - 5.

Kolmogorov–Smirnov tests have shown the raw data to be significantly non-normal and so Spearman, as opposed to Pearson, bivariate correlation analysis was undertaken to determine the level of correlation between the ADMs used and a measure of site size for each university. Site size was initially taken to be the number of static pages in a university web site. It is recognised the other statistical measures have been successfully used to show significant correlation with ADMs, particularly research activity indicators involving research ratings and/or staff numbers for individual universities. This approach is seen to have some advantage over the use of the number of pages (or links) in a university site as there is an element of normalisation. Larger universities would naturally be expected to have more pages in their web site and attract more inlinks and so any bivariate correlation could be explained by both variables being related to university size. After normalising for size, another explanation must be sought for any significant correlation found. However, there are few countries outside the UK for which there is a research assessment exercise sufficiently authoritative to be definitive for this purpose, although Australia are due to carry out a similar exercise, known as the Research Quality Framework (RQF) in 2008. However, as no research related metric common to all three academic webs could be identified, it was felt that using the number of static pages as a measure of site size would prove sufficient for purely descriptive purposes. Also, as UK RAE data was only available for the year 2001, it was felt that the use of page count data, which are available for each year, would give more accurate results than the use of static RAE data as a measure of university site size. Page count data is also used as a measure of university site size for the Webometrics Ranking of World Universities (www.webometrics.info).

Many webometric studies use inlink counts in preference to outlink counts as it is often argued that inlinks are more useful as indicators than outlinks, as outlinks are under the control of the site owners whereas inlinks are not. An additional technical problem with site outlink counts is that they depend upon a single site crawl, and are therefore more liable to crawler coverage problems than inlink counts, which are totalled from a number of different crawls. In order to avoid the problems associated with the used of either inlink or outlink counts in its comparison of ADMs, this study uses both.

6.4 Results

The inlink and outlink page, directory, domain and site ADMs output by SocSciBot tools, together with site size (measured as the number of static pages for each individual university and calculated using bespoke Visual Basic statistical analysis programs) were subjected to Spearman correlation testing using SPSS 12.0.

Correlation is a bivariate measure of the strength of the relationship between two variables. It varies from 0 (random relationship) to 1 (perfect linear relationship) or -1 (perfect negative linear relationship). Values close to 1 or -1 have high correlation while values close to 0 have low correlation. Spearman's rank correlation coefficient, denoted by the Greek letter ρ (rho), is a non-parametric measure of correlation, i.e. it assesses how well an arbitrary monotonic (changing in one direction only) function could describe the relationship between two variables without making any assumptions about the frequency distribution of the variables.

In principle, ρ is simply a special case of the Pearson product-moment coefficient in which the data are converted to rankings before calculating the coefficient but, unlike the Pearson product-moment correlation coefficient, it does not require the assumption that the relationship between the variables is linear. The raw scores are converted to ranks, and the differences between the ranks of each observation on the two variables are calculated.

ρ is then given by:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

where d_i = the difference between each rank of corresponding values
and n = the number of pairs of values

However, it is important to remember that a correlation, even a very strong one, does not mean that we should immediately jump to conclusions about causation. We should always be aware that the correlation in itself is no proof of assertion.

SocSciBot Tools were used to create standardised file-based, directory-based, domain-based and site-based link structure files from the original University of Wolverhampton Academic Web Link Database Project link structure files. Note that only links between UK, Australian and New Zealand universities were included.

The inlink and outlink page, directory, domain and site ADMs, together with site size for the same year as the link data (as measured as the number of static pages / domains for each individual university and calculated using bespoke Visual Basic statistical analysis programs) were then subjected to Spearman correlation tests using SPSS 12.0 and the results are shown in the tables below. The variables used in the correlation testing are therefore individual university site size (as measured as the number of static pages for each individual university and calculated using bespoke software) and the number of inlinks (Table 1) and outlinks (Table 2) aggregated at the page, directory, domain and site ADM level.

Table 6.1 Spearman Correlation Coefficients for site size (number of static pages) against Page, Domain, Directory and Site ADMs for UK, Australian and New Zealand Universities, 2000 – 2006 using inlinks

Web Space	Date of Crawl	Page Inlinks	Directory Inlinks	Domain Inlinks	Site Inlinks
New Zealand (8 sites)	Jul 00	0.643	0.643	0.643	0.082
	Feb 02	0.548	0.429	0.762*	0.000
	Jan 03	0.714*	0.714*	0.762*	0.031
	Dec 03	0.786*	0.833*	0.524	0.078
	Jan 05	0.905**	0.905**	0.524	0.412
	Jan 06	0.857**	0.857**	0.714*	0.412
Australia (38 sites)	Jul 00	0.492**	0.473**	0.511**	0.045
	Jan 02	0.652**	0.724**	0.731**	0.711**
	Mar 03	0.616**	0.694**	0.682**	0.659**
	Feb 04	0.750**	0.767**	0.793**	0.650**
	Mar 05	0.762**	0.780**	0.732**	0.733**
	Apr 06	0.767**	0.802**	0.757**	0.691**
United Kingdom (108-125 sites)	Jul 00	0.426**	0.486**	0.537**	0.550**
	Jul 01	0.841**	0.860**	0.867**	0.769**
	Jul 02	0.807**	0.833**	0.856**	0.762**
	Jun 03	0.789**	0.862**	0.876**	0.845**
	Jun 04	0.871**	0.887**	0.877**	0.876**
	Jul 05	0.883**	0.886**	0.880**	0.823**

* Correlation is significant at the 0.05 level (2-tailed) ** Correlation is significant at the 0.01 level (2-tailed)

Table 6.2 Spearman Correlation Coefficients for site size (number of static pages) against Page, Domain, Directory and Site ADMs for UK, Australian and New Zealand Universities, 2000 – 2006 using outlinks

Web Space	Date of Crawl	Page Outlinks	Directory Outlinks	Domain Outlinks	Site Outlinks
New Zealand (8 sites)	Jul 00	0.082	0.082	0.082	0.082
	Feb 02	0.690	0.810*	0.905**	0.655
	Jan 03	0.786*	0.810*	0.905**	0.764*
	Dec 03	0.905**	0.881**	0.810*	0.655
	Jan 05	0.833*	0.929**	0.738*	0.577
	Jan 06	0.690	0.643	0.743*	0.577
Australia (38 sites)	Jul 00	0.594**	0.601**	0.650**	0.545**
	Jan 02	0.754**	0.872**	0.884**	0.402*
	Mar 03	0.747**	0.754**	0.778**	0.319
	Feb 04	0.778**	0.798**	0.837**	0.323*
	Mar 05	0.831**	0.865**	0.853**	0.437**
	Apr 06	0.817**	0.873**	0.883**	0.255
United Kingdom (108-125 sites)	Jul 00	0.722**	0.756**	0.762**	0.689**
	Jul 01	0.860**	0.892**	0.893**	0.652**
	Jul 02	0.859**	0.876**	0.880**	0.660**
	Jun 03	0.840**	0.899**	0.910**	0.787**
	Jun 04	0.922**	0.925**	0.930**	0.816**
	Jul 05	0.919**	0.921**	0.919**	0.758**

* Correlation is significant at the 0.05 level (2-tailed) ** Correlation is significant at the 0.01 level (2-tailed)

The tables also show the levels of significance of the correlation. In this case, using the SPSS standard 2-tailed test showing significance at the 0.05 level (which means that the odds that the correlation is a chance occurrence are no more than 5 out of 100 i.e. a 95% chance that the correlation is not due to sampling variability) and the 0.01 level (which means that the odds that the correlation is a chance occurrence are no more than 1 out of 100 i.e. a 99% chance that the correlation is not due to sampling variability). The convention in most research is to use a significance level of 0.05 (a 95% confidence level) but both are shown here for comparison purposes.

6.5 Discussion

The results of Table 6.1 (inlinks) show that for the New Zealand academic space, the page and directory ADMs exhibit consistently strong correlations, followed closely by the domain ADM. The site ADM showed weak correlation, although this did appear to increase over time. The results for the Australian and UK academic web spaces show remarkable similarities. For these two countries, both the directory and domain ADMs exhibit similar, very strong correlations with site size followed closely by the page ADM and lastly the site ADM (although still showing significant correlation). It is also noticeable that the strength of the correlation for all four ADMs for the UK and Australian web spaces appears to be increasing over time.

In Table 6.2, the ADMs are compiled using outlinks, as opposed to inlinks. For the New Zealand academic web space, the directory and domain ADMs exhibit similar, very strong levels of correlation. The page ADM shows slightly less strong correlations with the site ADM correlation even weaker (although the use of outlinks shows a substantial increase in correlation levels when compared to the use of inlinks, shown in Table 6.1). Again, the UK and Australian academic webs show many similarities; although the domain ADM is shown to be the slightly more reliable, the directory and page ADMs also show very strong levels of correlation, with the site ADM again proving to be the weakest.

Quantitatively, the ADMs showing the *average* strongest and weakest correlations for each academic web space are displayed in Table 6.3 below:

Table 6.3 The Strongest and Weakest Spearman's Correlation Coefficients for ADMs using Inlinks and Outlinks for UK, Australian and New Zealand Universities (Page Site Size)

	Correlation (Inlinks)		Correlation (Outlinks)	
Academic Web Space	Strongest	Weakest	Strongest	Weakest
New Zealand	Page ADM (0.742)	Site ADM (0.169)	Domain ADM (0.697)	Site ADM (0.552)
Australia	Directory ADM (0.707)	Site ADM (0.582)	Domain ADM (0.814)	Site ADM (0.380)
United Kingdom	Domain ADM (0.816)	Site ADM (0.769)	Domain ADM (0.882)	Site ADM (0.727)

It may seem natural to draw the following conclusions from Table 6.3:

- The site (or university) ADM, when examined over time, shows the weakest levels of correlation against a standard measure of university site size and so would provide the least reliable measure across all three academic web spaces.
- The domain ADM appears to be the strongest general measure across the three academic web spaces over time (especially when used with outlinks).
- The use of outlinks, rather than inlinks, appears to show higher levels of correlation across the three web spaces over the time period in question.

However, while the conclusions above might appear obvious, the implicit assumptions that web page counts are correct and reliable, and that ideally all link measures should correlate perfectly with page counts should be questioned. For example, sites which have large numbers of library or other pages which do not contain any outlinks would be 'penalised' in any correlation analysis.

Taking the ADM that appears to give the strongest levels of correlation over the six year period, the following logarithmic graphs show the number of domain ADMs, using outlinks, against UK university site size from 2000 to 2005.

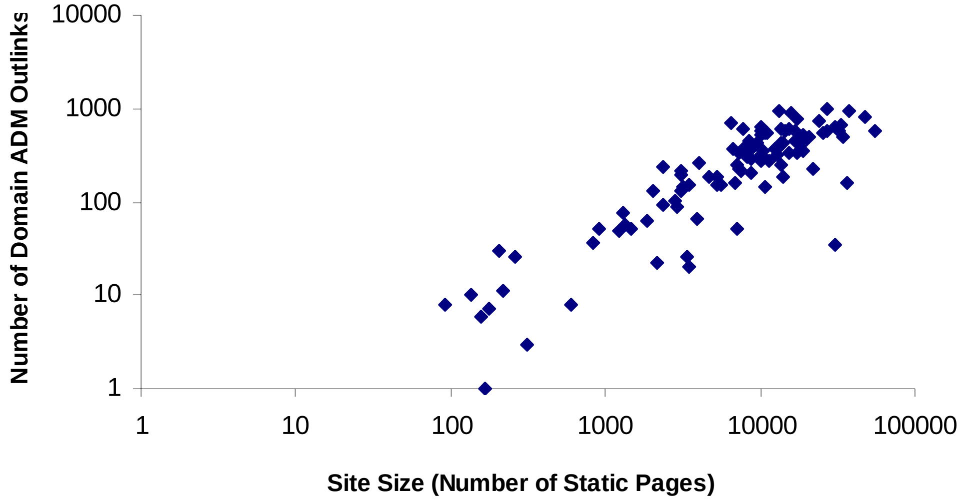

Figure 6.1 Number of Domain ADMs against Site Size
(for UK Universities in the year 2000)

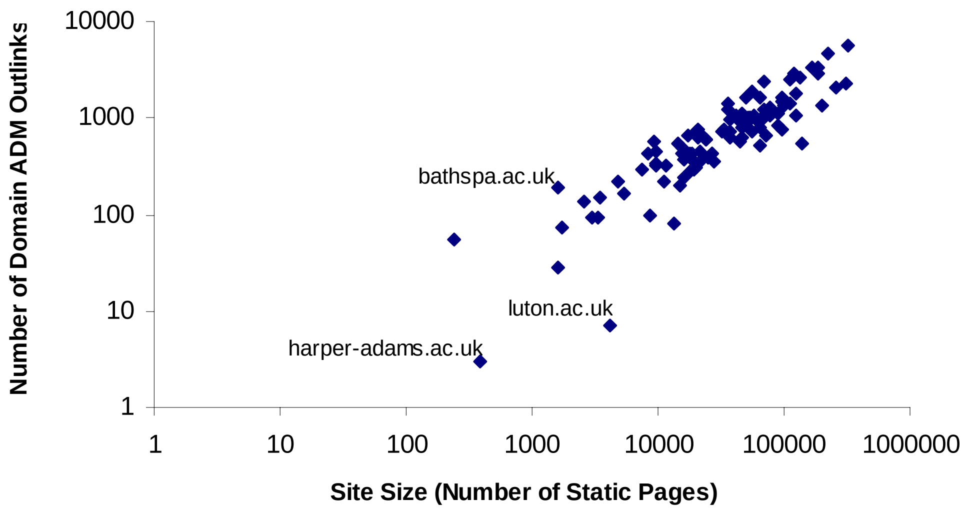

Figure 6.2 Number of Domain ADMs against Site Size
(for UK Universities in the year 2001)

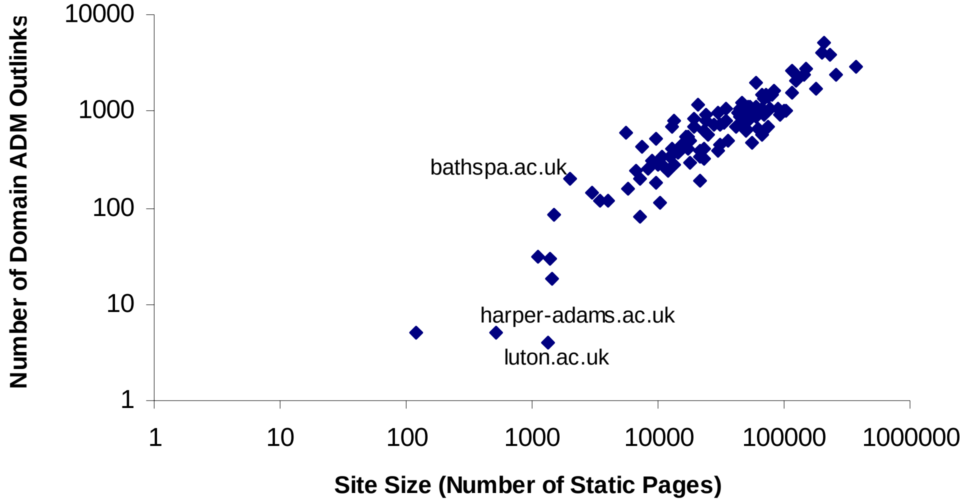

Figure 6.3 Number of Domain ADMs against Site Size (for UK Universities in the year 2002)

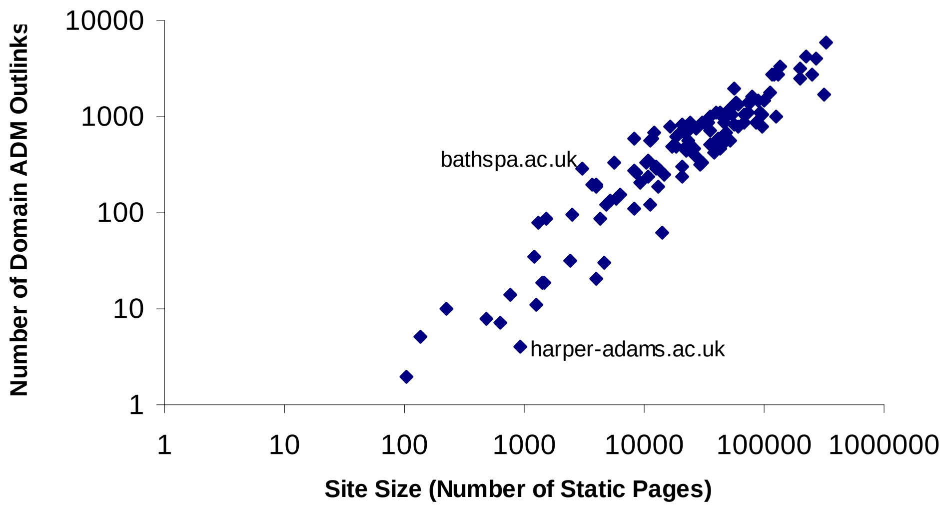

Figure 6.4 Number of Domain ADMs against Site Size (for UK Universities in the year 2003)

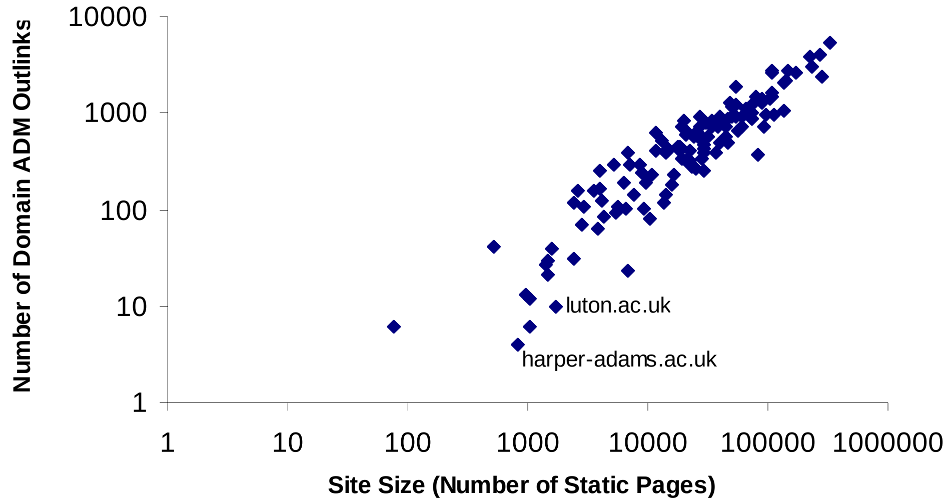

Figure 6.5 Number of Domain ADMs against Site Size (for UK Universities in the year 2004)

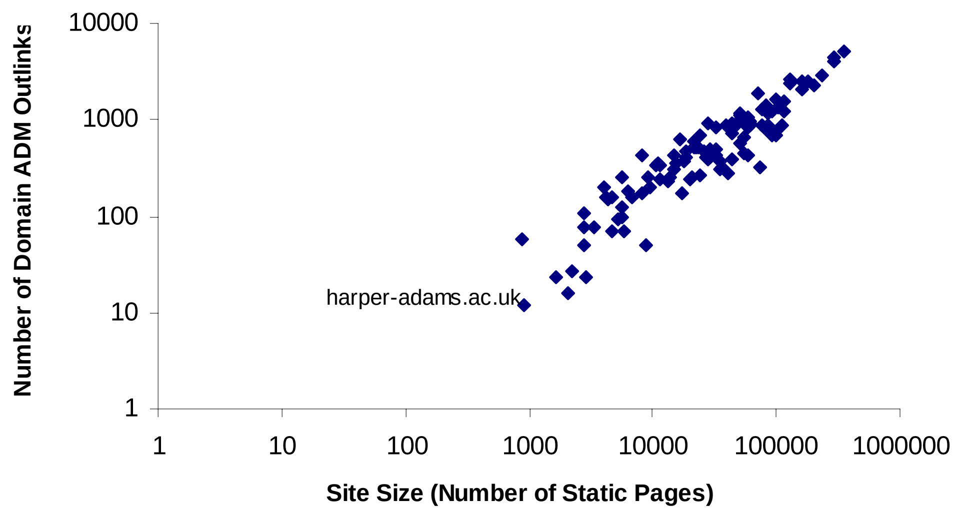

Figure 6.6 Number of Domain ADMs against Site Size (for UK Universities in the year 2005)

These graphs show the consistently high level of correlation between domain ADMs and university site size, and also identify the outliers for the respective data sets. Harper-Adams Agricultural College, Bath Spa University College and the University of Luton feature regularly as anomalies in these graphs as they have a disproportionate amount of domain ADM outlinks to the number of static pages in the university web site (or vice versa). The Harper-Adams Agricultural College link structure shows numerous links to university pages containing images while the University of Luton web site contains large numbers of links to its online journals and books. These are all examples of university web sites which contain large numbers of pages but, as most of the links are internal site links, they do not contain a proportionate number of outlinks according to the graphs, and so show up as outliers. Bath Spa University College web site shows links to a large number of personal web sites and this university is an example of a site having a large number of outlinks proportional to the number of static pages.

These outliers are examples of individual universities exhibiting behaviour very different to the overall trend and show that the correlation graphs appear not to be the aggregation of all universities showing similar behaviour but the aggregation of different behaviours.

It should also be noted that the use of Spearman's correlation coefficient is considered to be unreliable when used with very small sample sizes and, with only 8 universities, the New Zealand academic web space may not be sufficiently large to give truly accurate statistical results. The results in Tables 6.1 and 6.2 may support this conclusion as only 18.75% of the correlations from the New Zealand academic web space are significant at the 0.01 level. Comparing this with the fact that 89.6% of the correlations from the Australian academic web space (which has 38 universities) are significant at the 0.01 level while 100% of the correlations from the UK academic web space (which has between 108 and 125 universities depending on the year) are significant at the 0.01 level would suggest that a larger sample size may increase the accuracy of the correlation calculations.

If, as suggested, New Zealand correlations are taken to be unreliable and are removed from the study, the results do not significantly change. For the UK and Australian academic webs alone, the domain ADM displays a slightly higher correlation than the directory ADM for both outlinks (0.848 and 0.836 respectively) and inlinks (0.758 and 0.754 respectively). The page ADM shows less (although still significant) correlation at 0.803 for outlinks and 0.721 for inlinks while the site ADM shows the weakest correlations for both outlinks (0.554) and inlinks (0.676).

It was noted as an area of concern that there may be an issue in the use of the number of static pages as a measure of site size as any subsequent comparison with the page ADM could be expected to produce high levels of correlation. In an attempt to allay these concerns, the inlink and outlink page, directory, domain and site ADMs were again subjected to Spearman correlation tests, this time using the number of source domains as a measure of site size for individual universities. The result of this analysis is displayed in Tables 6.4 – 6.6 below.

Table 6.4 Spearman Correlation Coefficients for site size (number of source domains) against Page, Domain, Directory and Site ADMs for UK, Australian and New Zealand Universities, 2000 – 2006 using inlinks

Web Space	Date of Crawl	Page Inlinks	Directory Inlinks	Domain Inlinks	Site Inlinks
New Zealand (8 sites)	Jul 00	-0.577	-0.577	-0.577	-1.000**
	Feb 02	0.524	0.357	0.786*	0.000
	Jan 03	0.381	0.381	0.810*	0.187
	Dec 03	0.429	0.381	0.881**	0.078
	Jan 05	0.311	0.311	0.790*	0.249
	Jan 06	0.359	0.359	0.814*	0.332
Australia (38 sites)	Jul 00	0.480**	0.517**	0.576**	-0.076
	Jan 02	0.678**	0.732**	0.836**	0.714**
	Mar 03	0.818**	0.835**	0.911**	0.636**
	Feb 04	0.880**	0.852**	0.880**	0.668**
	Mar 05	0.852**	0.843**	0.865**	0.612**
	Apr 06	0.880**	0.852**	0.877**	0.574**
United Kingdom (108-125 sites)	Jul 00	0.521**	0.596**	0.699**	0.593**
	Jul 01	0.764**	0.778**	0.822**	0.667**
	Jul 02	0.725**	0.747**	0.818**	0.675**
	Jun 03	0.689**	0.755**	0.820**	0.726**
	Jun 04	0.820**	0.817**	0.844**	0.779**
	Jul 05	0.751**	0.752**	0.781**	0.651**

* Correlation is significant at the 0.05 level (2-tailed) ** Correlation is significant at the 0.01 level (2-tailed)

Table 6.5 Spearman Correlation Coefficients for site size (number of source domains) against Page, Domain, Directory and Site ADMs for UK, Australian and New Zealand Universities, 2000 – 2006 using outlinks

Web Space	Date of Crawl	Page Outlinks	Directory Outlinks	Domain Outlinks	Site Outlinks
New Zealand (8 sites)	Jul 00	1.000**	1.000**	1.000**	1.000**
	Feb 02	0.643	0.810*	0.952**	0.655
	Jan 03	0.548	0.619	0.810*	0.546
	Dec 03	0.548	0.643	0.833*	0.655
	Jan 05	0.299	0.359	0.886**	0.581
	Jan 06	0.491	0.587	0.916**	0.581
Australia (38 sites)	Jul 00	0.717**	0.779**	0.872**	0.637**
	Jan 02	0.475**	0.720**	0.884**	0.316
	Mar 03	0.641**	0.797**	0.862**	0.272
	Feb 04	0.608**	0.752**	0.830**	0.294
	Mar 05	0.671**	0.735**	0.804**	0.303
	Apr 06	0.704**	0.759**	0.852**	0.167
United Kingdom (108-125 sites)	Jul 00	0.741**	0.787**	0.894**	0.749**
	Jul 01	0.776**	0.790**	0.845**	0.532**
	Jul 02	0.712**	0.746**	0.820**	0.521**
	Jun 03	0.711**	0.781**	0.853**	0.704**
	Jun 04	0.817**	0.830**	0.861**	0.713**
	Jul 05	0.799**	0.795**	0.841**	0.648**

* Correlation is significant at the 0.05 level (2-tailed) ** Correlation is significant at the 0.01 level (2-tailed)

Table 6.6 The Strongest and Weakest Spearman's Correlation Coefficients for ADMs using Inlinks and Outlinks for UK, Australian and New Zealand Universities (Domain Site Size)

	Correlation (Inlinks)		Correlation (Outlinks)	
Academic Web Space	Strongest	Weakest	Strongest	Weakest
New Zealand	Domain ADM (0.584)	Site ADM (-0.026)	Domain ADM (0.900)	Page ADM (0.588)
Australia	Domain ADM (0.824)	Site ADM (0.521)	Domain ADM (0.851)	Site ADM (0.332)
United Kingdom	Domain ADM (0.797)	Site ADM (0.682)	Domain ADM (0.852)	Site ADM (0.645)

The results using the number of domains as a measure of site size given in Tables 6.4 – 6.6 above confirms the findings from Tables 6.1 – 6.3 when the number of pages was used as the site size measure. The tables show that the domain ADM consistently shows the highest average level of correlation for all three academic webs. This may be expected when using the number of domains as a measure for site size but the fact that it reflects the results established when the number of pages was used as a measure of site size reinforces these conclusions.

Tables 6.4 and 6.5 also demonstrate similarities between the levels of correlation found for the UK and Australian academic webs and that the use of outlinks appears to produce higher levels of correlation than the use of inlinks. After the domain ADM, the directory ADM shows the next highest correlations, followed by the page ADM. The site ADM once again appears to be the least reliable model, showing the lowest levels of correlation in five of the six results.

6.6 Conclusions

During previous ADM-based studies, there has been much discussion about the relative accuracy and reliability of the various ADMs. The domain and directory models appear to be the most successful in reducing the impact of anomalous linking behaviour between pairs of websites, with the directory-based URL counting model being better for analysing interlinking between universities, at least in the UK (Thelwall, 2002d; Thelwall and Wilkinson, 2003b; Payne and Thelwall 2004). However, the domain ADM would also appear to be a good choice, with Thelwall (2004b) finding that the standard domain ADM emerged as the logical choice for comparison purposes when counting links from UK, Australian and New Zealand universities. Thelwall and Harries (2004a) analysed UK inter-university link count data using two different ADMs, the domain ADM and the page ADM, arguing that the domain ADM was the more valid model because of the known link anomalies between UK universities.

The consensus seems to be that the most reliable ADMs have been the directory and domain models, although the hybrid range models have also had some success (Thelwall and Wilkinson, 2003b). However, given the lack of a universal best choice of ADM, another option would be to use several. Thelwall and Vaughan (2004a) introduced several new versions of the PageRank algorithm using ADMs, but with inconclusive results. The new approach seemed to work well for sets which included pages from other web sites but did not work well in ranking pages from the same site.

Although correlation between the number of pages in a site, and the number of inlinks / outlinks to and from this site is not a new finding, this study has shown that, while strong correlations do indeed exist between pages and links across all three academic web spaces over the last six years, consistently higher results are obtained when the web links are aggregated at the directory, and especially domain level. High correlations are evidence of a more consistent model, using the assumption that the number of links to and from a web site should be proportional to the site, other factors being equal.

The domain ADM has an additional advantage in that the Top Level Domain (TLD) names are relatively simple to extract from hyperlinks and automatically process. However, the success of the domain model is not conclusive enough to be able to claim that it is the sole definitive model for link analysis research, although the results do show that it is significantly better than the default page model. Additionally, examination of the outliers shows individual universities exhibiting behaviour very different to the overall trend. This suggests that the strong levels of correlation observed may not be due to the aggregation of all universities showing similar trends but to the aggregation of partly different trends.

Conversely, aggregating at the site (or university) level appears to conclusively provide less reliable results than using the page as the standard unit of measure, and this finding appears to hold true over all three academic webs and for each time period analysed over the last six years. The fact that the results from Tables 6.4 – 6.6 (where the number of domains was used as a measure of site size), are in line with the findings from Tables 6.1 – 6.3 (where the number of pages was used as a site size measure), shows that these results are consistent for differing measures of university site size.

These results can be partially explained by remembering that the use of ADMs will balance anomaly elimination with loss of data. A higher level of aggregation will eliminate link anomalies, but will also lose data due to the aggregation. For this study, the directory and domain models have emerged as the most suitable for producing high levels of correlation, and it may be that the page ADM does not sufficiently address the anomalies while the site ADM, due to the high level of aggregation, loses too much data.

While previous specific studies have shown that more accurate results can be obtained through the use of the domain or directory ADM, this study has shown that these results are consistent over time, and across different countries' academic web spaces and, although it seems likely that these results would be generally applicable, it is not inconceivable that there would be countries to which they would not apply, for example if a URL, directory or domain structure was used that was substantially different to that used in the UK, Australia or New Zealand. That said, the very high correlation found for domain and directory ADMs for the UK, Australian and New

Zealand academic webs does encourage the belief that they may well be robust enough to withstand transportation to other countries. In any case, these positive results appear to strengthen the case for using web link analysis as a tool with the potential to reveal underlying trends in academic website interlinking.

7 Do Academic Link Types Change over Time?

7.1 Introduction

Several previous studies have used web links, either solely or as part of a more integrated solution, to classify web pages. Furnkranz (1998) used web links to classify pages using the assumption that it was easier to classify web pages using information on pages that point to a page instead of using information provided on the page itself. Calado et al., (2005) evaluated how the link structure of the web can be used to determine a measure of similarity appropriate for document classification. Tests performed on a web directory showed that link information alone allows classification of documents with an average precision of 86% and that when combined with a traditional text-based classifier, precision increased to values of up to 90%. Koehler (1999b) offers three separate approaches to aid the classification of web documents including the use of URL characteristics, whereas Haas and Grams (1998) believed that classification systems for both web pages and web links would be most effective if they were developed in tandem. They proposed an integrated classification system based on a content analysis of 75 web pages and the 1500 web links they contained. Harrison (2002) explored the semantic and rhetorical principles underlying link creation, proposing a classification of links according to their primary function while Crowston and Williams (2000) used a random sample of 1000 URLs from the AltaVista search engine to classify around 50 different genres on the web.

Link creation motivation studies are vital for developing an understanding of how link counts should be interpreted and often include some form of page classification exercise. For example, Park (2002) conducted a survey of 64 Korean webmasters to assess their motivation for linking to other web sites. He found that, although web links were generally created for either navigational functionality or business purposes, webmasters require that the credibility of hyperlinked web sites be higher than average when deciding to hyperlink to them. Chu (2005) analysed a random sample of links from academic institution's web sites and generated a list of reasons for hyperlinking. On the whole, almost 50% of all the inlinks examined were created for pointing to resource or directory information provided at the target web sites. In addition, 73% of all the inlinked sites analysed were linked to for reasons relating to service or home page while less than one third (27%) of the links were made out of research or teaching/learning motivations. Park, Barnett and Nam (2002) regarded the number of inlinks to a web site as an indicator of site credibility, finding that site inlink counts were significantly related to visiting behaviour and perceived web site credibility. Kim (2000) investigated motivations for creating links in electronic publications in order to examine the analogy between citations and scholarly linking in electronic journals. Scholarly, social and technological reasons were identified, with most links having more than two reasons for creation.

From an academic perspective, a number of studies have attempted to classify university web pages. Thelwall (2001d) carried out a classification of pages in an attempt to differentiate between research oriented and non-research oriented links. A later study of the top 100 most highly linked-to pages in UK universities found no recreational pages in the data set, instead finding it dominated by university home pages (Thelwall, 2002i). Thelwall, Harries and Wilkinson (2003) took a random sample of 586 pairs of interlinked academic web sites and classified them using the 68 subject categories of the UK's RAE. The results were compared to the number of active researchers in each subject area in order to determine which subjects had large or small web presences for their size.

Thelwall (2003c) took a sample of 100 random inter-site links to UK university home pages and proposed four new types of motivation; ownership, social, general navigational and gratuitous. A similar study was undertaken by Kousha and Horri (2004) for Iranian universities finding that, of the 440 links studied, 63% were made for gratuitous or navigational reasons. Links between universities have also been analysed with a three-faceted (mathematics, physics and sociology) link source and target categorization scheme (Harries et al., 2004).

The most detailed academic interlinking motivation studies so far involve source and target page and link classification exercises using different categories for link creation motivation. Wilkinson et al., (2003) took a random collection of 414 links between UK academic institutions, downloaded both the source and the target pages and classified them according to the apparent motivation for their creation. They found that, by combining similar categories, more reliable ones were formed and that, although less than 1% of hyperlinks targeted formal scholarly publications such as journal articles or conference papers, over 90% of targeted material was in some way related to research or other scholarly activity. Bar-Ilan's (2004c) academic link studies included categories for the type of source and target page of inter-university links in Israel. Although the categories and data collection methods used by these two studies are different (a specialist web crawler and a commercial engine respectively), the results are similar. Bar-Ilan's (2004c) 'research-related' category contains 20% of all links studied while 27% of links in the Wilkinson et al., (2003) classification fell into the 'research support and resources', 'research partners' or 'research reference' classes. Excluding 'superficial' and 'technical' links from Bar-Ilan's (2004c) study shows that approximately 86% of links are related to scholarly activity, reflecting the findings of Wilkinson et al., (2003).

7.2 Research Question

This paper is a first attempt to address the important issue of how different types of academic web links change over time. It combines web-based longitudinal data (e.g. Koehler 1999b; 2002; 2004, Fetterly et al., 2003) with a link classification exercise (e.g. Bar-Ilan 2004b; 2004c, Wilkinson et al., 2003) to answer the following research question:

How and why does the distribution of types of academic web links change over time?

It is important to identify the types of links between university web spaces. The extent to which inter-university web links correlate to research-related activities (and in particular, to the equivalent of journal citations) suggests the level of success which could be expected when applying standard bibliometric techniques to academic web links.

7.3 Methods

A specialist webometrics crawler (Thelwall, 2001a) initially collected the raw data in the form of text files containing the hyperlink structure of each UK, Australian and New Zealand university. These text files are publicly available as part of the University of Wolverhampton Academic Web Link Database Project (Thelwall, 2002/3) and several academic web link analysis papers have already been published using this data. Information science software (SocSciBot Tools, socscibot.wlv.ac.uk) was then used to generate random links, together with their associated source pages, from these text files. These links were selected in such a way that approximately the same number were taken from each university's site and the raw data used during the link classification exercise is available at Appendices 6 - 11.

When attempting a classification exercise of this type, the research can be carried out by either examining the web links and pages in an attempt to assign reasons for link creation or by interviewing the web page authors to ascertain their motivation. In this study, the former option was chosen. If the primary purpose of this study was to identify the motivation behind the creation of the links studied, then questioning the authors directly may have provided information of a more authoritative nature. However, as the focus on link types for the current study is topical rather than motivational, it was felt that individual researchers categorising the links based on pre-determined criteria was acceptable. Although it could be argued that the reason for the creation of a link would naturally dictate the type of that link and that an aspect of motivational study is implicit in this research, this study does not attempt to directly address the question of the motivation behind the creation of the links but instead attempts to group the links using defined topics or categories with a view to quantitatively establishing how each link category changes over time.

Kousha and Horri (2004) utilised the academic link classification scheme proposed by Wilkinson et al., (2003) but this study uses instead the scheme proposed by Bar-Ilan (2004b) in which she categorised Israeli inter-university links for their apparent intentions. As discussed above, the classification process is very subjective and the more general categories used by Bar-Ilan (2004b) were chosen in an attempt to increase the reliability of the classification exercise. Also, as noted in Thelwall (2004c), to try to ensure high levels of inter-indexer consistency, the categories should be kept as simple as possible and categories irrelevant to the research goals should be avoided.

When reporting results, care should be taken to explain the categories used as there may be a high degree of ambiguity in some category descriptions. As well as the authors' interpretations of the Bar-Ilan (2004b) descriptions, exemplars for each category are shown in Table 7.1.

Table 7.1 Link Type Descriptions and Exemplars

Technical	Links to pages on which technical services were the main theme e.g. online libraries, databases or software applications
Exemplars	University list of bibliographic databases → BookFind online database Information page for maths lecturers → Numerical and statistical library software
Research oriented	Links to pages containing specific research activities e.g. abstracts of academic papers or presentations, lists of resources on a specific research topic, pages detailing scientific activity and conferences, homepages of researchers which include lists of publications, participation in research projects
Exemplars	Lecturers homepage → Journal paper Medical ethics reading list → Biomedical ethics readings
Professional	Links to pages on which most of the information is related to non-academic, non-recreational work-related activities e.g. company and professional body homepages or general business-related material
Exemplars:	Project Management links page → Project Management Exhibition web page List of professional bodies in the Leisure and Tourism Industry → Tourist Society page
Educational	General university education-related, as opposed to research-related, links and targeted mainly at student activity e.g. links to taught courses, course lists, collaborative working
Exemplars	University library web page → Guide to referencing styles Guide to collaborative learning → Student activity web page
Administrative	Purely academic organisational links e.g. to university, research group or departmental homepages
Exemplars	IT resource page → University home page List of educational publications → The Centre for Asia Pacific Social Transformation Studies
Personal	Links to personal homepages where the professional side is not emphasised or personal bookmarks where the majority of links do not seem to be professionally oriented
Exemplars	List of fiction authors → Authors personal home page Building and construction materials assessment page → Lecturers personal home page
Social/Leisure	Links to pages related to hobbies and leisurely interests e.g. social groups, travel, list of restaurants, links to friends or other general interest web sites
Exemplars	Scottish Tourism page → Volunteers travel organisation home page University 'City Life' page → Guide to local restaurants
Superficial	Embedded links that are inserted as part of the template for the page or as part of the code used to build the page e.g. a link from a logo or a verifier link
Exemplars	Programming course page → Java link embedded in logo Departmental publication page → Web page template stat counter
Navigational	Organisational links purely intended to aid navigation e.g. site maps and links to another part of the same page
Exemplars	List of law courses → Different part of same page Information for international students → Image map (site map)
Informative	Specific, specialised informational links that occur too infrequently to belong to any of the previous categories e.g. an error message
Exemplars	Network server status page → Error message Health and Safety bulletin → Notification of agenda
Other	Links of a specific type not covered by any of the values above, which appeared too rarely to be classified in a specific category e.g. a memoriam page
Exemplars	Web server usage page → Web stats generating company

Note that some of the studies mentioned (e.g. Thelwall, 2001a; 2002; 2003, Thelwall & Harries, 2003) concentrated on page type classification while in others (Bar-Ilan, 2004a; 2004b) the links themselves were classified. The current research is interested only in the classification of web links, as this is part of a wider study into academic web linking. The classifiers in this study examined the source page and target page in order to generate more information and thus allow the link itself to be classified more accurately.

Although the original data set used in the study was collected at random, initial investigations showed that many of the pages and links have now changed or disappeared, especially for the pages in the academic web dating from the year 2000. A decision was thus made to use only the first randomly selected 100 links from each country's academic web which were verifiable using the Internet Archive. The Internet Archive is becoming an increasingly important tool for conducting longitudinal studies of the Internet maintaining a record of the evolution of the web and is a key resource for webometricians. The Internet Archive Wayback Machine records all the dates on which a page was found, keeping copies of every indexed version so that old versions can still be examined.

7.3.1 Pilot Study

This study attempts to incorporate elements from an information science approach to link analysis (Thelwall, 2004c) and, as such, begins with a pilot study. The purpose of this study is to assess whether there will be a sufficient number of links to produce significant results, and, if so, whether the types of links found are broadly consistent with the research goals. A pilot study can save a lot of wasted effort if a project is unsuitable because of the quantity or quality of the links found.

The pilot study visited a sample of 60 randomly selected links (10 from the New Zealand, Australian and UK academic webs for each of the years 2000 and 2005/6) to assess their type and determine how well they fit the goals of the research. The links visited were chosen to be as random as possible, and classified from a visit to the source page and also to the target page, based on the Bar-Ilan (2004b) schema. The results of the pilot study are shown in Table 7.2 below:

Table 7.2 Results of Pilot Link Classification Exercise

Type of Link	Number of Occurrences
Professional	8
Research Oriented	11
Educational	8
Administrative	9
Informative	0
Personal	6
Social/leisure	2
Technical	10
Navigational	2
Superficial	4
Other	0
Total	60

Although no attempt to introduce any longitudinal aspect was made at this stage, the pilot study seems to suggest that the types of links present are of sufficient quality and quantity to warrant further study.

7.3.2 Full-Scale Random Sampling

Following the methodology outlined by Thelwall (2004c), in order to support the validity of the results reported, a classification of 600 random (but Internet Archive indexed) links were taken from the full data set. This number was considered sufficient to generate a high enough proportion of links relevant to the research question. The sampling of links used the random link generator of SocSciBot Tools (based upon a random number generator function) which takes an equal number of links from each university and ensures that all links from the same university have an equal chance of being selected. The breakdown of the 600 links is given in Table 7.3.

Table 7.3 Breakdown of 600 Randomly Selected Links

Academic Web	Crawl Dates	Number of Occurrences
UK	June-July 2000	100
New Zealand	July-August 2000	100
Australia	July-August 2000	100
UK	July 2005	100
New Zealand	January 2006	100
Australia	April 2006	100

Classification exercises are inherently difficult because web pages are known to not conform to existing genres particularly well (Crowston & Williams, 2000). Single researcher classifications are problematic as they are based on one person's perception and interpretation of link creation motivations. A lack of cross-checking by additional classifiers was noted as a limitation in the Kousha and Horri (2004) study and meant that their results could not be generalised to other areas of research.

However, most multi-researcher studies have also found links very difficult to classify in the sense of not getting high levels of agreement between classifiers despite the use of detailed classification schemes (Harries et al., 2004). In some cases, despite the fact that the classification scheme had been jointly devised and tested by the researchers, there was still a considerable level of disagreement in the results.

Link creation motivation appears to be a subjective issue with numerous opinions for the reasons behind link creation making classifications of pages and links genuinely difficult to produce. This seems to be a problem endemic to researcher-based page classification exercises and this paper was no exception, i.e. the classification was problematic initially due to a low level of agreement on categories by independent researchers. This study chose an approach which involved an additional researcher classifying 10% (i.e. 60) randomly selected links. Following extensive discussion, a consensus was reached and a subsequent inter-indexer consistency test showed 82% agreement (i.e. 49 of the 60 links) between the two researchers, which is significantly high to establish some validity for the original researcher's classifications.

Previous research has shown that web link-based studies should not only consider link counts, but also the motivations for linking, in order to ensure the validity of such research. However, the highlighted difficulties in classifying link motivations and interpreting link counts appears to be a major problem for all web link-related studies.

7.4 Results

The results from the full-scale random sampling exercise are shown in Table 7.4 with the most popular link types first:

Table 7.4 Results of Full-Scale Link Classification Exercise

	Number of Links					
Type of Links	UK 2000	NZ 2000	AU 2000	UK 2005	NZ 2006	AU 2006
Technical	24	25	22	13	15	12
Research Oriented	13	14	14	24	22	24
Professional	20	16	20	17	19	18
Educational	14	10	11	11	13	13
Administrative	10	12	8	9	8	10
Personal	14	16	14	3	3	7
Social/leisure	0	1	2	9	9	5
Superficial	0	2	2	7	7	6
Navigational	4	2	3	3	2	2
Informative	1	1	3	2	1	3
Other	0	1	1	2	1	0
Total	100	100	100	100	100	100

UK: United Kingdom, NZ: New Zealand, AU: Australia

From Table 7.4, there appears to be no significant difference in the number of links between the three academic webs for each category for both the 2000 and 2005/6 time periods. However, to show this conclusively, a Chi-Square test of independence was used to test the association between two categorical variables; the number of links per academic web and the link type.

For the purpose of the test, Table 7.4 is split into two 3 x 11 tables, one for 2000 and another for 2005/6. The degrees of freedom (d.f.) in a test of independence are equal to (number of rows – 1) x (number of columns -1). An Excel spreadsheet was then used to calculate the Chi-Square value and the level of significance.

In a test of independence the null and alternative hypotheses are:

Ho: The two categorical variables are independent.
Ha: The two categorical variables are related.

The Chi-Square test shows here that number of different types of links in the UK, New Zealand and Australian academic webs are not significantly different for 2000 (Chi-Square = 9.8087, d.f. = 20, significance = 0.9715) or for 2005/6 (Chi-Square = 8.2524, d.f. = 20, significance = 0.9901). This shows that there is no evidence of a difference between the distribution of types of link between the three countries, either in 2000 or in 2005/6, and we can therefore reject the null hypothesis in both cases.

7.5 Discussion

Before examining the longitudinal aspect of the results, some general trends can be initially identified. For example, 77% of all links fall into either the ‘technical’, ‘research oriented’, ‘professional’, ‘educational’ or ‘administrative’ categories i.e. they have a role in research or education. This is supported by the fact that pages that could be classed as recreational (i.e. ‘personal’ and’ social/leisure’) account for only 14% of all links. The remaining 9% of links (‘superficial’, ‘navigational’, ‘informative’ and ‘other’) are general non-academic, non-recreational links.

These findings support those of Wilkinson et al., (2003) and Bar-Ilan (2004b) who found that the majority of pages (probably at least 86% in both the UK and Israel) play some role in research or education, but fall short of formal scholarly communication, and that recreational pages form only 9% of academic links in the UK while social and superficial academic links in Israel total 13%.

A longitudinal analysis of the results revealed some interesting patterns and Table 7.5 below aggregates the three academic webs and summarises the number of links for the years 2000 and 2005/6:

Table 7.5 Longitudinal Summary of Full-Scale Link Classification Exercise

	Number of Links	
Type of Links	2000	2005/6
Technical	71	40
Research Oriented	41	70
Professional	56	54
Educational	35	37
Administrative	30	27
Personal	44	13
Social/leisure	3	23
Superficial	4	20
Navigational	9	7
Informative	5	6
Other	2	3
Total	300	300

For Table 7.5 above, Chi-Square = 59.936, d.f. = 10 and significance = 0.0000. This demonstrates that there is a statistically significant difference between the distribution of types of links in 2000 and 2005/6. However, further examination of Table 7.5 would suggest that there are no discernable trends evident for the ‘professional’, ‘educational’, ‘administrative’, ‘navigational’, ‘informative’ or ‘other’ categories between 2000 and 2005/6. Carrying out a Chi-Square test for independence on these six categories alone confirms this (Chi-Square = 0.7576, d.f. = 5, significance = 0.9797).

Nevertheless, the ‘technical’, ‘research oriented’, ‘personal’, ‘social/leisure’ and ‘superficial’ categories do display noticeable differences over time (Chi-Square = 59.123, d.f. = 4, significance = 0.0000). The fact that different types of web links behave differently over time supports similar findings with regard to web pages (Koehler, 2002). The substantial increase in the ‘superficial’ category may be explained by the recent wide-spread proliferation of hyperlinks within logos, either to other departments within the university or external companies. For example, none of the ‘superficial’ links in 2000 were logo hyperlinks whereas by 2005/6, this had risen to 13 of the 20 links identified.

Also, the significant decrease in ‘personal’ pages may be attributable to another recent phenomenon; the popularity of social networking sites and blogs. These sites make it easier to host personal information in dedicated, collaborative online environments and are more suitable than university web sites for storing and sharing personal information.

It may be the case that while personal content has been moved to dedicated ‘Web 2.0’ sites, university web sites have become recognised as the most appropriate place to host social content for academic groups rather than individuals. As university web sites mature, it may be accepted that there is a place for recreational, as well as educational content and this may be a contributory factor in the noted increase in the number of ‘social/leisure’ links observed. Of the 23 links classified as ‘social/leisure’ in 2005/6, 11 were dedicated to travel (7 for travel abroad and 4 to local tourist attractions), 4 to sporting activities, 3 to local guides such as restaurants or cinemas and 5 were classed as miscellaneous.

More difficult to explain is the decrease in the number in 'technical' links between the years 2000 and 2006. One would expect that the number of links to online libraries, databases, journals and applications to have increased over time but this does not appear to be the case. One possible explanation, borne out by a previous study is that the number of actual technical links has remained the same but, because the number of source pages of universities within the three academic webs has experienced a marked increase between 2000 and 2002 (Payne & Thelwall, 2007a), as a percentage of overall links, 'technical' links may appear to have decreased.

Another explanation, supported by the fact that 45% (32 of the 71) of 'technical' links in 2000 but only 28% (11 out of 40) in 2005/6 were to online libraries or databases is that certain online repositories now offer the same content only previously available by linking to a collection of others and so fewer technical links are now needed to access the same content.

The increase in 'research oriented' links, from 14% of all links in 2000 to 23% in 2005/6 was expected. As more universities are forming online collaborations and are reacting to calls to develop online, open access repositories of research material, the number of 'research oriented' links could be expected to increase. The figure of 23% is greater than the 20% 'research-oriented' figure produced by Bar-Ilan (2004c) and, although the number of links classified here are lower, this may provide preliminary evidence that universities are taking increasing advantage of research-based web resources.

As one of the largest categories, and one critical to this body of research, all 'research oriented' links were revisited, and the number of links which could be directly compared to journal citations was calculated (Table 7.6).

Table 7.6 Links Equivalent to Citations in the Six Data Sets.

	Number of Links					
Type of Links	UK 2000	NZ 2000	AU 2000	UK 2005	NZ 2006	AU 2006
Research Oriented	1	2	0	4	2	6

UK: United Kingdom, NZ: New Zealand, AU: Australia

These results further reinforce the findings above that universities appear to be taking advantage of research-related online resources and may be evidence of an increased use in online research repositories. The results above also support the findings of Wilkinson et al., (2003). They found that less than 1% of hyperlinks targeted formal scholarly publications such as journal articles or conference papers and therefore could be considered to be the equivalent of formal citations. This study shows that in the year 2000, 3 out of 300 links (1%) were the equivalent of journal citations but by 2005/6, this had grown to 12 out of 300 links (4%). Although this suggests that the number of academic links directly comparable to journal citations is increasing, further research would be needed to corroborate this.

It is possible that certain types of links historically found in static pages are now more often found in dynamic pages and therefore concealed from the majority of web crawlers. For example, in core university web sites, general administrative and prospectus information now tends to be displayed using dynamic pages. This may offer a partial explanation for some of the changes in link types although at present there is no evidence to support this.

It was expected that this link classification exercise would highlight some inter-country differences but examination of the results shows no identifiable disparity between the link types of the UK, New Zealand and Australian academic webs. Previous studies have identified differences between these academic webs, particularly for New Zealand but, in this case, all three webs appear to display the same characteristics, even when viewed from a longitudinal perspective.

7.6 Conclusions

This research addresses concerns expressed in Bar-Ilan (2004b) over only considering links between the academic institutions of one country by considering the links for academic webs within three countries and, while there are no notable trends apparent with regard to link types across the three academic web spaces of the UK, New Zealand and Australia, there are some major differences in the way that certain types of academic links change over time. Significant increases in the 'research oriented', 'social/leisure' and 'superficial' categories were identified as well as notable decreases in the 'technical' and 'personal' categories. Some of these differences can be explained by general changes in the management of university web sites i.e. 'research-oriented' and 'social/leisure' links and some by more wide-spread Internet trends i.e. 'personal' and 'superficial'. The decline in 'technical' links over the years is surprising and may or may not be attributable to the explanations suggested above. Further study would be needed to more clearly identify the reason for this.

Also worthy of further research is the apparent increase in citation-equivalent links, although the numbers are very low. The analogy between journal citations and hyperlinks is at the very heart of many webometric studies and, if this increase continues, there may eventually be enough to have a significant overall influence on university hyperlinking.

Academic web links appear to represent a wide range of activities, covering research and educational, as well as recreational aspects. Links are an important part of research and education and, although they do not appear to be the direct equivalent of journal citations, this study suggests that this may be changing in terms of the increasing proportion of research-related links.

8 A Longitudinal Study of Academic Webs: Growth and Stabilisation

8.1 Introduction

Many webometric studies have counted pages or links from the web sites of the universities in one or more countries, identifying differences between universities and countries (Thelwall, Vaughan & Björneborn, 2005). However, a major problem endemic to all web link analyses is that, because the web is continuously evolving, any web study may be obsolete by the time it is published in the academic literature (Thelwall & Payne, 2005). Changes in size may be important in absolute terms or in relative terms, with relative differences being particularly important if there are non-uniform growth rates. For example, if the university web sites in country A are doubling in size annually whereas those in country B are static, then their relative sizes will change quickly over time and any statistics that compare site sizes or links between the two will be misleading after a short period of time. If the web sites in both countries are either static or growing at the same linear rate, then the relative size difference will be constant, and so comparative statistics would be valid for a longer period of time. Hence it is very important to know how web link analysis results vary over time, with a low rate of variation lengthening the shelf-life of webometric results.

Many information scientists have come to realise the importance of longitudinal studies while conducting their own research. According to Rousseau (1999), 'collecting time series should be an essential part of Internet research'. However, although there have been numerous hyperlink studies of national and international academic institutions, very little research has been carried out from a purely longitudinal perspective. It would be natural to assume that the web is continuing to grow exponentially, but it is also possible that parts of the web have reached a saturation point, with web site sizes stabilising.

8.2 Research Question

The aim of this chapter is to identify and track changes in three academic webs over time (UK, Australia and New Zealand), tracking aspects of academic webs including site size and overall linking characteristics, and to provide theoretical explanations of the changes found. The choice of these three countries is driven solely by the availability of historical data. They are similar countries in the sense of being English-speaking, richer nations and part of the Commonwealth (thecommonwealth.org).

Specifically, this chapter addresses the following research question:

What is the trend over time for the average web site size, and average inlink count, of UK, Australian and New Zealand academic webs?

8.3 Methods

The raw data used during this study was derived from the link structure text files of the universities of the UK, Australia and New Zealand over a five-year period beginning in July 2000. These text files were obtained as part of the on-going University of Wolverhampton Academic Web Link Database Project (chapter 4.3).

The university text files were then processed using a suite of specially written programs designed to work with the structure of the text files produced by the crawler, sorting, counting and analysing the link data, before being subjected to further data analysis, and displayed graphically. The data used to produce Figures 8.1 – 8.6 can be found in Appendices 12 – 14.

While other longitudinal studies such as Koehler (1999b) and Fetterly et al., (2003) used random samples of web pages, this approach was not required in this case as a complete data set was available from the University of Wolverhampton Academic Web Link Database Project (Thelwall, 2002/3). Additionally, initial research revealed clear patterns in the data so that the use of advanced aggregation methods such as ADMs (Thelwall, 2002e) was considered unnecessary.

8.4 Results

The first results to be extracted from the database files concentrated on using charts to show changes in the three academic webs in question. Average web site size graphs were produced for UK, Australian and New Zealand universities over the five-year period (Figure 8.1). Site size in this case is measured by the number of static pages within each individual university site. Note that the data was crawled at five specific points over the time frame in question and this is reflected in the shape of the graph. The graph is satisfactory for identifying trends over time, but does not show fluctuations between the recorded data points.

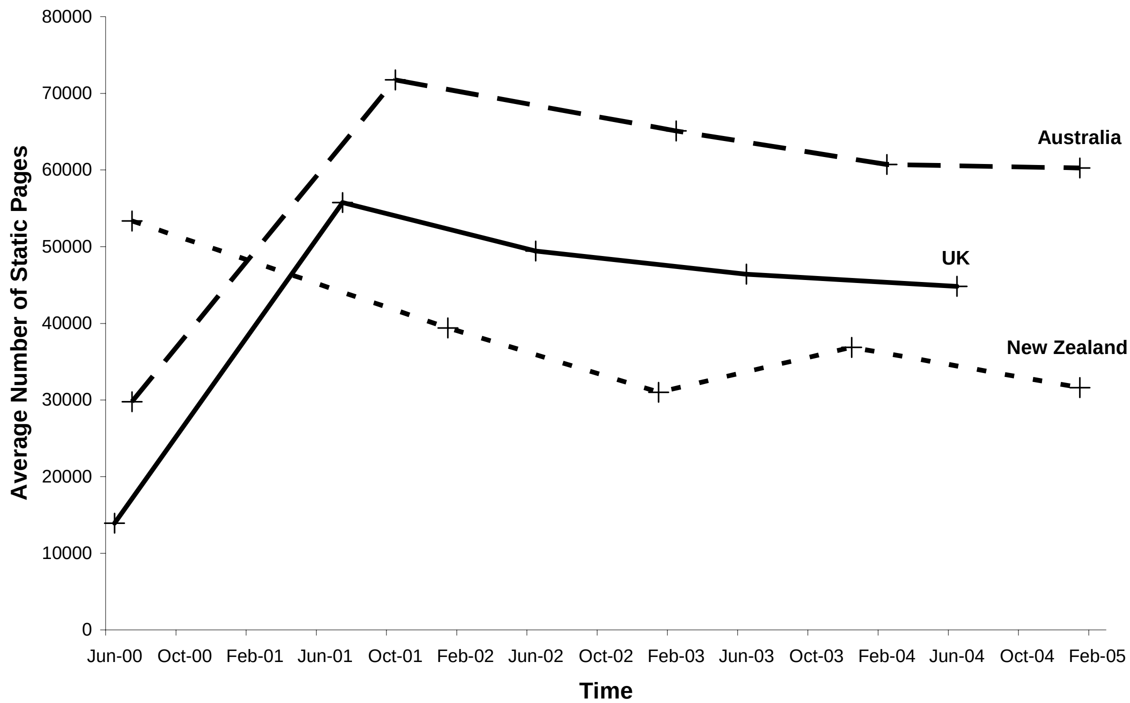

Figure 8.1 Average Number of Static Pages of UK, Australian and New Zealand Universities against Time

From Figure 8.1, the average web site sizes of the UK and Australian universities exhibit similar trends of rapid growth followed by slight shrinking, with the turning point being in 2001-2002. The change in site size between the two webs stayed within a range of 1686 pages across the entire time period. Both academic webs experienced an increase in average site size of almost 42,000 pages during the initial time frame before recording more modest consecutive falls over the next three time periods.

The graph of the average site size for the New Zealand universities in Figure 8.1 displays different characteristics. While the UK and Australian universities have experienced a net increase in average site size over the five-year period, the average New Zealand university web site size decreased, experiencing only one period of moderate growth over this time period.

Table 8.1 Change in Average Web Site Size for UK, Australian and New Zealand Universities over Time

	Average Web Site Size								
	June 00 –May 01	% change	June 01 –May 02	% change	June 02 –May 03	% change	June 03 –May 04	% change	June 04 – May 05
UK	13930	300.19	55747	-11.3	49445	-6.12	46419	-3.45	44816
Australia	29783	140.91	71749	-9.85	64680	-6.1	60735	-0.77	60269
New Zealand	53362	-25.18	39393	-21.3	31004	18.88	36858	-16.25	30870

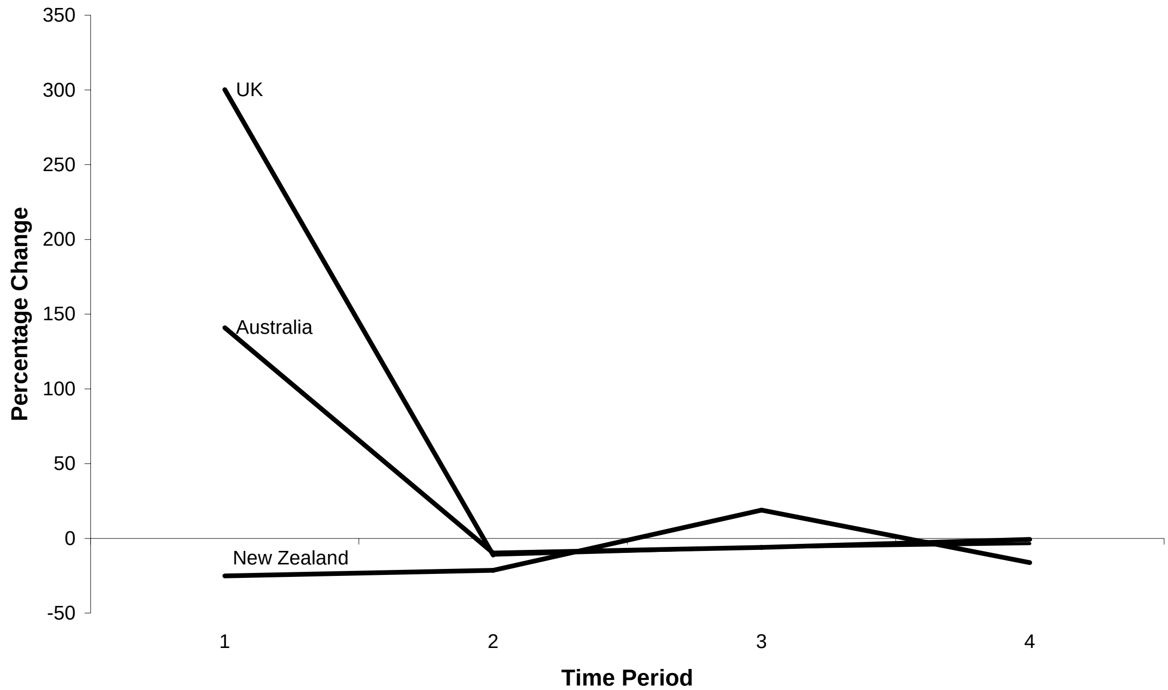

Figure 8.2 Percentage change in average web site size for UK, Australian and New Zealand universities against Time

Plotting the percentage change between consecutive data points from Table 8.1 gives the graph shown in Figure 8.2 above. This clearly shows that, in absolute terms, the amount of change experienced by the average web site for the three academic webs decreases in each consecutive time period, and this appears to be consistent for UK, Australian and New Zealand universities alike. This shows that the change to the average academic web site size is decreasing, which means that the average university site size for the UK, Australia and New Zealand has become more stable over a five-year period.

Figure 8.3 below shows the average number of static pages for all UK universities along with the average number of static pages for the University of Sussex, Queens University Belfast, the Open University and Brunel University. These four universities were chosen as examples of individual UK universities exhibiting behaviour very similar to the overall UK average, with the Open University apparently displaying the closest similarities.

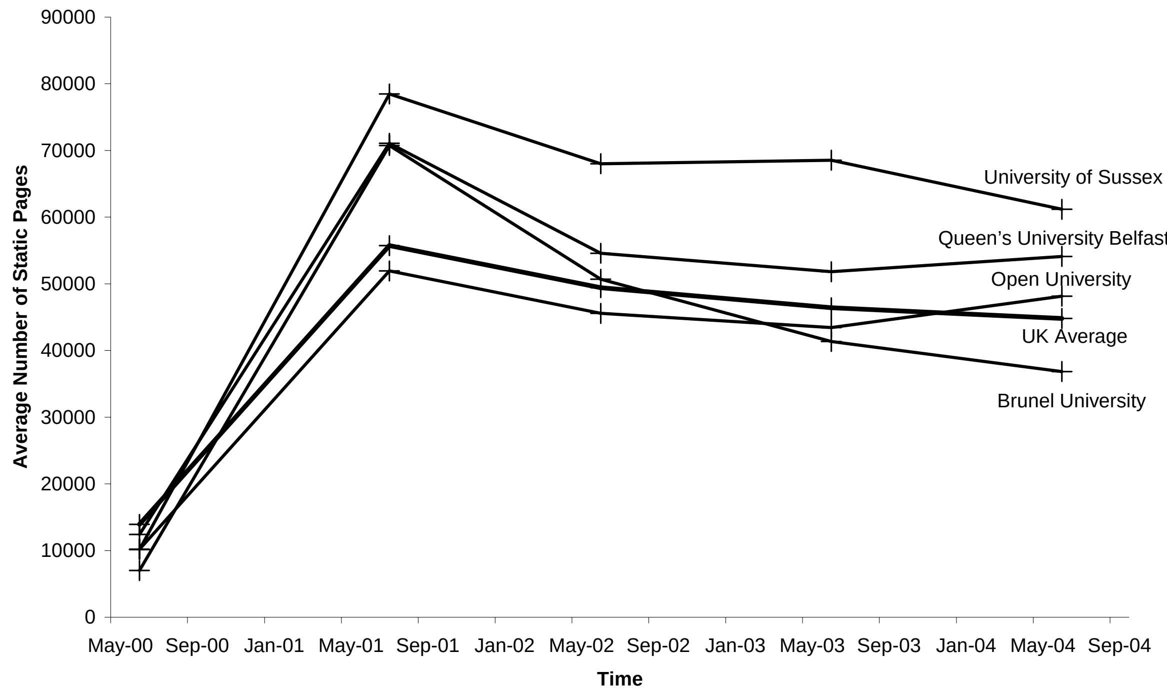

Figure 8.3 Average Number of Static Pages for Four UK universities with Average Trends and the UK Average against Time

Figure 8.4 below shows the average number of static pages for all UK universities along with the average number of static pages for the University of Edinburgh, University of Aberdeen, University of Wales, Lampeter and the University of Buckingham. These four universities are examples of individual UK universities exhibiting behaviour very different to the overall UK average.

This shows that the graph of the UK average is not the aggregation of all UK universities showing similar trends but is the aggregation of partly different trends. Indeed, only 5 of the 83 UK Universities (Brunel University, University of Derby, Kingston University, Queen Mary, University of London and the University of Strathclyde) exhibited the same growth, decline, decline, decline pattern exhibited by the UK average.

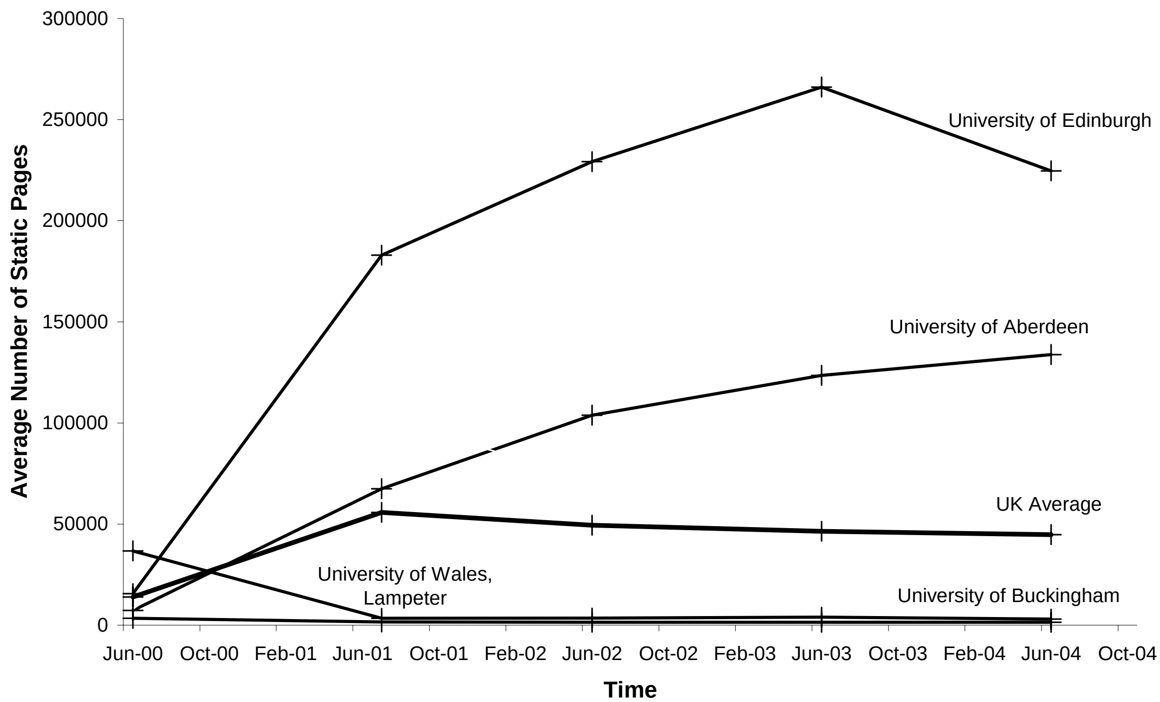

Figure 8.4 Average Number of Static Pages for Four UK Universities with Non-Average Trends and the UK Average against Time

Further examination of Figure 8.1 raised the possibility that the overall shape of the UK average graph may be the combination of two distinctly different graphs for older and newer UK universities, and that the initial sharp increase displayed during the year 2000 may be due to the new universities increasing their web presence. However, Figure 8.5 below shows this not to be the case. While it is apparent that the graphs for the average number of static pages for old and new universities (as defined by Hahessey et al., 2002) are distinctly different, the shape of the graphs exhibit similar periods of growth and decline. It simply appears to be the case that the older universities have had a longer period of time to build up more of a web presence, and this is reflected by the greater number of average static pages in their sites.

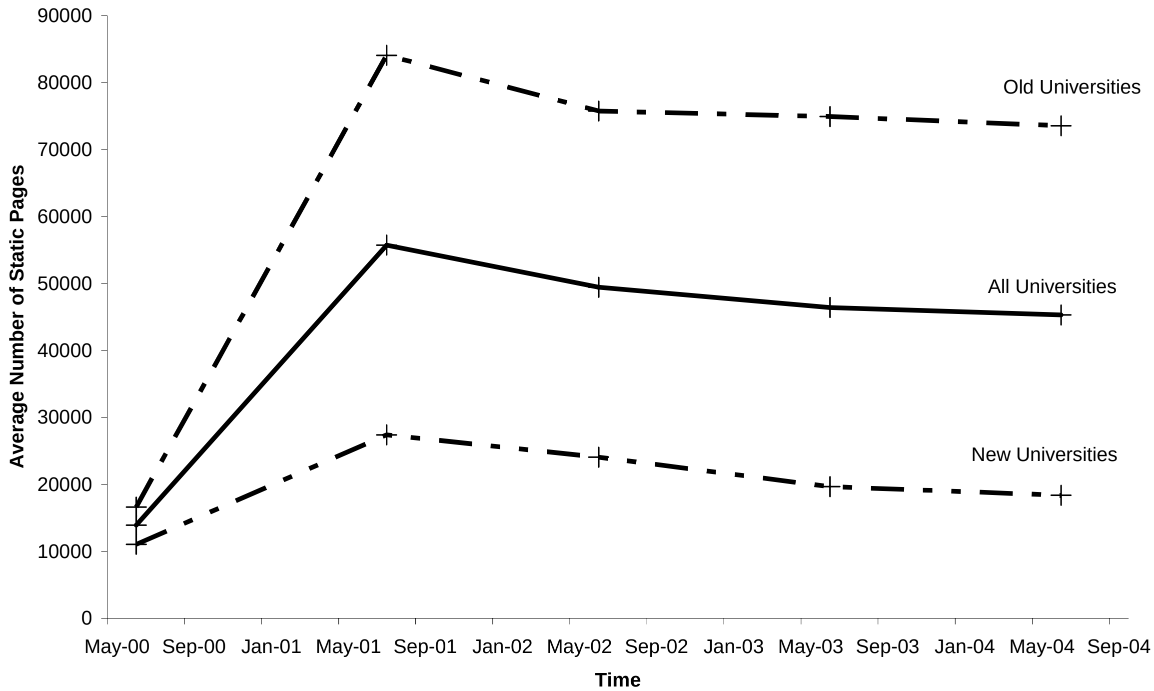

Figure 8.5 Average Number of Static Pages for New, All and Old UK Universities against Time

While Figure 8.1 shows the change in the average number of static pages of UK, Australian and New Zealand universities, Figure 8.6 below shows the change in the median number of static pages.

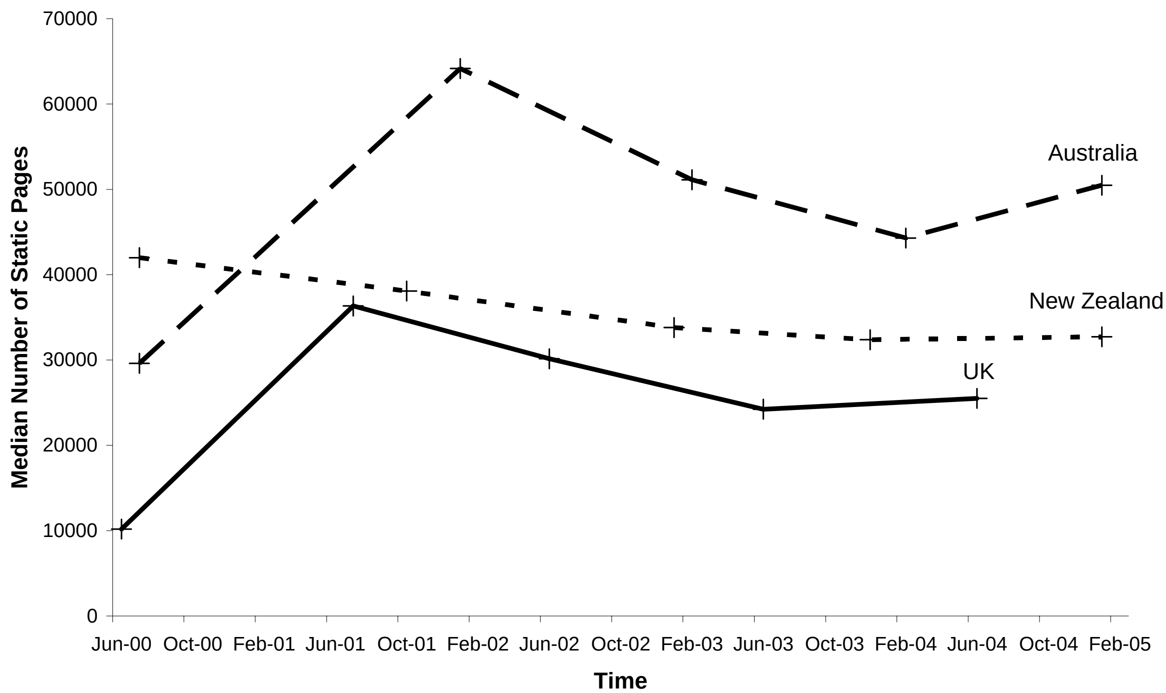

Figure 8.6 Median Number of Static Pages of UK, Australian and New Zealand Universities against Time

Again, the graphs for the UK and Australia exhibit striking similarities, while the graph for the New Zealand universities shows a marked difference. In this case, the decrease in the median of the number of static pages for the New Zealand universities is even more apparent.

The change in the average and median of the number of static pages between consecutive data points from Figures 8.1 and 8.6 shows that, in absolute terms, the amount of change experienced by the average web site for the three academic webs decreases in each consecutive time period, and this appears to be consistent for UK, Australian and New Zealand universities. This shows that the change to the average academic web site size is decreasing, which means that the average university site size for the UK, Australia and New Zealand is tending to become more stable over the five-year period.

While the results from Figures 8.1 and 8.6 appear to show that the change in site size for the three academic webs in question is decreasing over a five-year period (i.e. they are no longer expanding), it was hypothesised that this apparent change may be due to the fact that universities are increasingly taking advantage of newer web technologies which produce embedded links not counted by the specialist information science crawler used. With this in mind, Figure 8.7 below shows the average number of dynamic pages (defined as having a '?' in the URL signifying that these pages were generated by web applications utilising dynamic scripting languages such as ASP and PHP), and Figure 8.8 shows the number of non-HTML documents (such as Microsoft Word documents and Adobe PDF documents), detected by the crawler as being pages linked to by static pages over the five-year time frame. The raw data used to construct Figures 8.7 and 8.8 can be found in Appendices 15 – 20.

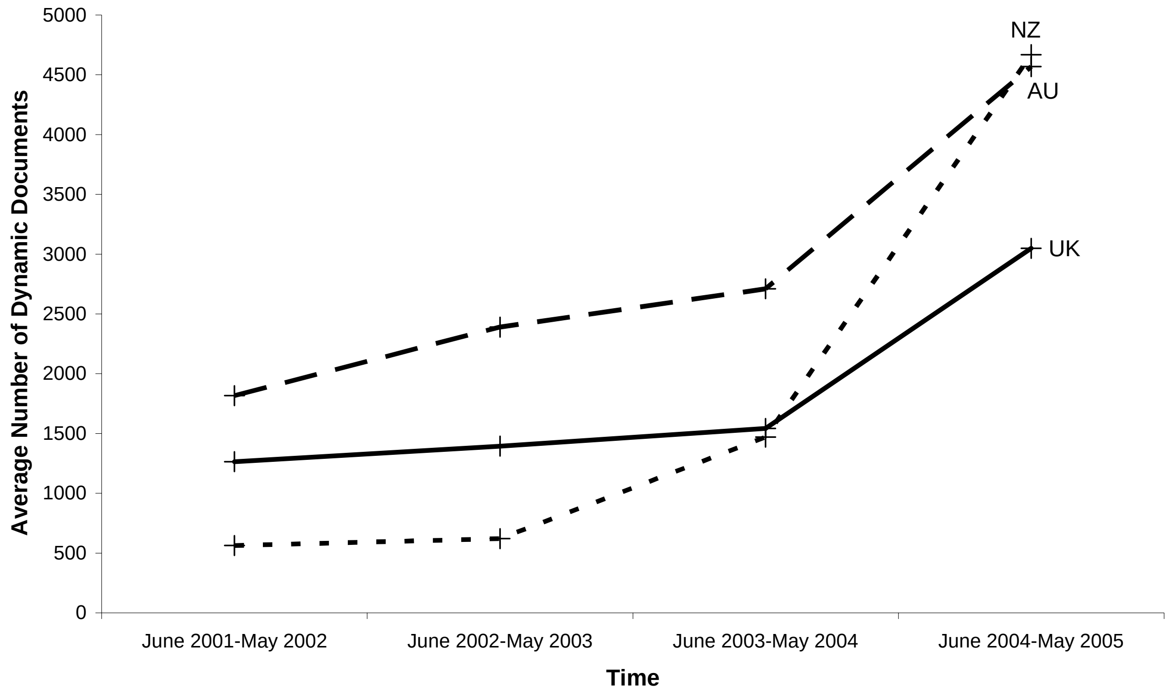

Figure 8.7 Average Number of Identified Dynamically Generated Pages for UK, Australian and New Zealand Academic Webs against Time

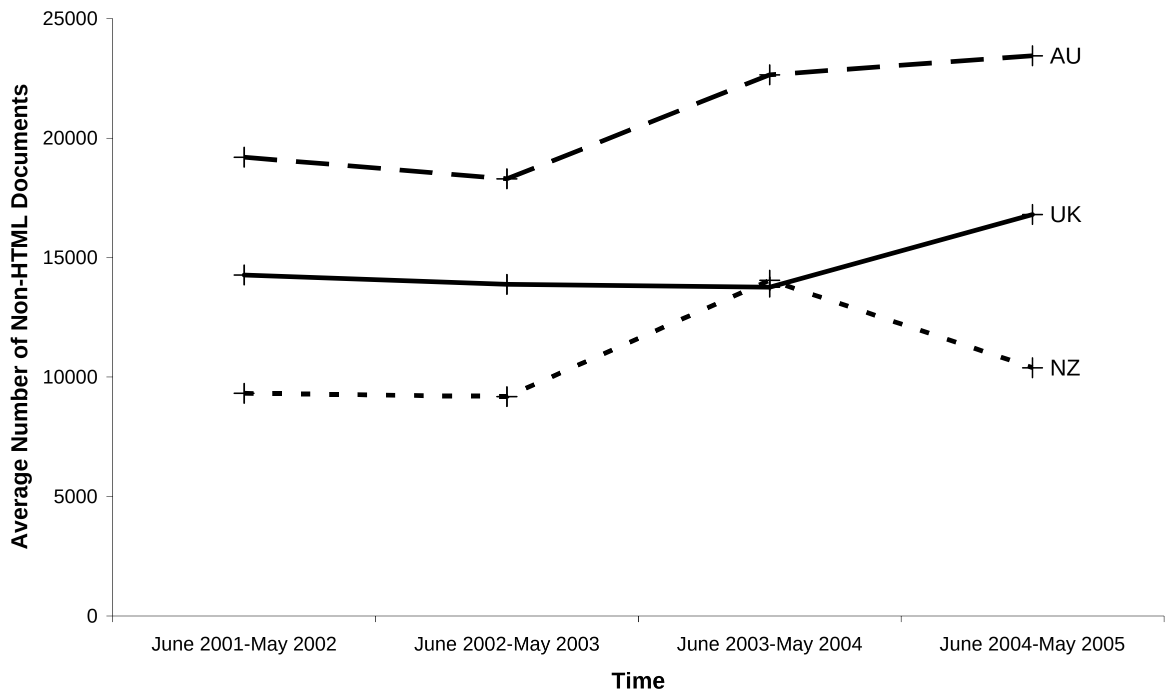

Figure 8.8 Average Number of Non-HTML Documents in UK, Australian and New Zealand Academic Webs against Time

Figure 8.7 above shows that the average number of identified dynamic pages has at least doubled over the time period in question for all three academic webs. New Zealand shows the greatest increase with 4106 pages, compared to an increase of 2750 pages for Australian universities and 1785 pages for UK universities. Figure 8.8 shows an increase in the number of non-HTML documents for the period June 2001 to May 2005 of 2533 for the UK, 4252 for Australia and 1064 for New Zealand universities.

Although all three academic webs show an overall increase in the number of identified dynamically generated pages and the number of non-HTML documents used, which does suggest a definite trend towards the uptake of advanced web technologies producing documents and dynamically generated pages from which the crawler could not extract HTML links, the numbers involved would not significantly affect the shape of the graphs in Figure 8.1 and therefore do not impact on the findings of this research.

Figure 8.9 below shows the average number of links to universities in each of the three academic webs (site inlinks) against time for the five-year period of the study. The raw data used to construct this graph can be found at Appendices 21 – 23. This graph shows links from other universities in the same country, excluding links between different pages of the same university and non university links from the same country. It should also be noted that the UK average graph has omitted the outlying data for the London Metropolitan University, which was created in August 2002 by merging the University of North London and London Guildhall University.

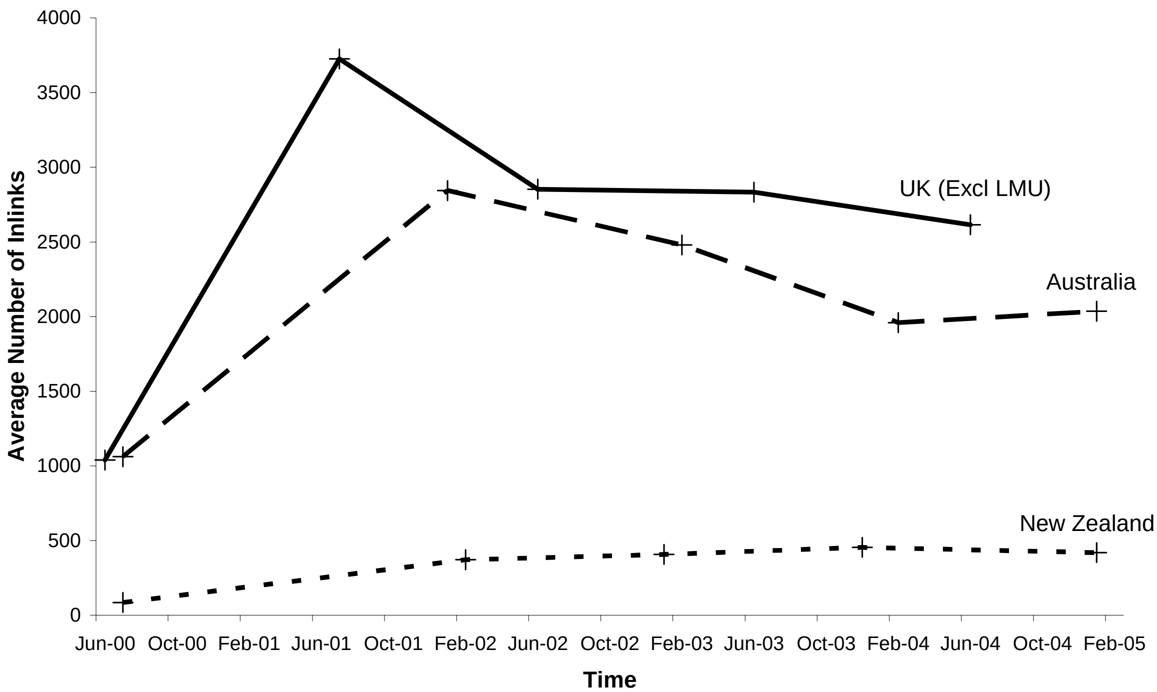

Figure 8.9 Average Number of Inlinks to UK, Australian and New Zealand Universities against Time

Again from Figure 8.9, there appear to be more similarities between the UK and Australia than New Zealand, but in this case it is clear from the graph that, as the average number of inlinks over time is higher at the end of the five-year time period than at the start, there is an upwards trend apparent in the number of inlinks for all three academic webs. The amount of change experienced by the average inlink count for the three academic webs in absolute terms does not decrease with each consecutive time period. This is in contrast to the average site size, shown in Figure 8.1, which appears to be stabilising.

While there may be a general trend towards stabilisation in the number of inlinks to UK, Australian and New Zealand universities in absolute terms, it is not as obvious as the corresponding trend for the number of static pages, and certainly not conclusive.

8.5 Discussion

This chapter first concentrated on identifying changes in the average web site size (as measured by the number of static pages in each site) for UK, Australia and New Zealand academic webs. It was found that the graphs for UK and Australian universities exhibit striking similarities, both in their shape and behaviour. The similarities between the two webs are all the more remarkable given that there exists no formalised recommendations for web page creation or guidance for official academic web sites in any individual country, let alone in two countries with such an obvious geographical disparity.

Examination of the graph of the average site size for the New Zealand academic web found it to exhibit different characteristics. This difference in behaviour may be explained by the fact that there is a much reduced sample size for New Zealand, i.e. there are only 8 New Zealand universities in the Academic Web Link Database, compared to around 110 for the UK and 38 for Australia. This makes it much easier for a New Zealand university with a large number of site web pages, such as the Victoria University of Wellington, to dominate the results. However, this may have an alternative explanation as Smith and Thelwall (2002) compared linking between UK, Australian and New Zealand universities and found that New Zealand was relatively isolated on the web, suggesting that the academic web in New Zealand is somewhat insular; relatively well interconnected but less well known internationally. This finding is also in line with the results of a previous bibliometric study for journals (Glänzel, 2001) which shows New Zealand to be one of the most isolated of the advanced nations in terms of international scientific co-authorship.

The differences for the New Zealand universities apparent in the graphs of Figures 8.1, 8.6 and 8.9 may be other examples of the insular behaviour of this academic web, particularly with regard to the graph of the number of inlinks shown in Figure 8.9.

Following on from the changes identified in Figure 8.1, the average number of static pages, the results were corroborated in Figure 8.6, the median number of static pages. Again, there were remarkable similarities between UK and Australian universities, with the graph of New Zealand universities exhibiting distinctly different characteristics.

It was hypothesised that the distinct shape of the average number of static pages for UK universities over time was a combination of dissimilar graphs for older and newer universities, with newer universities increasing their web presence and playing 'catch-up' to the older universities. Figure 8.5 shows this not to be the case; although the two graphs for older and newer universities show differences when plotted, the underlying shape of the graphs is the same, with both experiencing the same periods of growth and decline, albeit with a vastly different number of average static pages. The older universities have a far greater average number of static pages, and this could be explained by them having had a greater amount of time to establish their presence on the web and subsequently a greater number of static pages in their web sites.

Of real concern was the fact that the data collection method used (the specialist information science web crawler) by design only crawled static web links. It was recognised that the apparent stabilisation in the average number of static pages for university web sites may not be attributable to the fact that these web sites have stopped expanding, but to the fact that the universities in question are making use of new Internet technologies such as PDF documents or Active Server Pages which are capable of producing links undetectable by the web crawler used. If this were the case, it would undoubtedly call into question the validity of these results. Therefore, effort was made to identify and display the trends associated with the three academic web's use of dynamically generated web pages and non-HTML documents.

In the case of dynamically generated pages, Figure 8.7 shows clear evidence to suggest that universities are making substantial use of dynamic web page technology. However, due to the number of dynamically generated pages in question, these findings do not impact on the significance of this research. Figure 8.8 also appears to identify a definite upward trend with regard to the use of non-HTML documents, but this was not enough to have a significant impact on the result of Figure 8.1 due to the overall average number of static pages.

It may be the case that, if the data for Figure 8.1 was adjusted to include the results for both the average number of dynamically generated pages and the average number of non-HTML documents, this may result in a horizontal line from the second half of 2001 onwards i.e. the later portion of the graph exhibiting neither growth nor decline. This would still show a trend towards stabilisation but may be indicative of the latter part of an S-type or sigmoidal curve typical of limited population growth, as seen in plants and animals as well as at the molecular and cellular level. Due to the programmatic differences in producing the data for Figures 8.1, 8.7 and 8.8, no effort has been made to amalgamate the graphs, as it was felt that this had the potential to produce misleading results.

Figure 8.9 shows the average number of inlinks to the universities of the three academic webs. While this may display some evidence of a trend towards stabilisation, not one of the three countries in question displayed a clear, unbroken trend. This appears to be in contrast with the graphs of the average and median number of static pages shown in Figures 8.1 and 8.6 respectively.

8.5.1 Outliers

Figure 8.9 has omitted the data referencing the London Metropolitan University as stated. This decision was taken once it was realised that the summation of the 'londonmet.ac.uk', 'lgu.ac.uk' and 'unl.ac.uk' domains over the 2001 – 2003 period while the merger between the University of North London and London Guildhall University was taking place had a disproportionate effect on the overall average, and consequently the shape of the graph. Care should always be taken to ensure that the removal of outliers does not simply become a convenient way to eliminate all awkward data points; in this case a theoretical justification is provided to support the removal decision to avoid the validity of the data being compromised.

8.5.2 Limitations

Although these results appear to suggest that the site size for the three academic webs in question is stabilising, there may be other explanations for the shape of these graphs. Chu, He and Thelwall (2002) state that webometric research must be conducted with caution as both the data source (in this case the hyperlink structure of the three academic webs) and data collection instrument (the specialised information science crawler) may have deficiencies.

In this chapter, while effort has been made to validate the findings by producing results accounting for embedded links and dynamically generated web pages, there may be a deficiency in the data collection instrument in its inability to count obscured links. Obscured links are URLs that can be accessed by web users in ways that are difficult or impossible for web crawlers. The link extractor part of a crawler is not capable of extracting all links from web pages because some can be stored in formats that are in practice impossible for them to decode. For example, with programs running through the web browser, such as JavaScript, Java, Shockwave and Flash, it is not possible to easily extract URLs since these may be built by the code itself when running. This would mean a site using this kind of technology without the back up of HTML links would not be covered completely, although in some universities, including the University of Wolverhampton, certain types of pages such as main university and departmental sites have reverted back to static.

The following are some examples of applications which make use of obscured links:

- Embedded programs are programs running through the web browser, such as JavaScript, Java, Shockwave and Flash. It is not possible to easily extract URLs from embedded programs since these may be built by the code itself when running. This would mean a site using this kind of technology without the back up of HTML links would not be covered completely.
- Automatically generated pages are web pages that are created in response to web surfers' actions and do not exist before they are requested. The web page created is a genuine web page with its own unique URL but if a crawler visited the search engine site again, it would not find the same page as it was created in response to the query and then effectively destroyed. There are several web technologies that make it easy to do this, including PHP (Hypertext Pre-processor) and Microsoft's Active Server Pages (ASP)
- Non-HTML documents types with a hyperlinking capability, for example online Microsoft Word, PDF documents or images. No links are ever extracted from non-HTML documents.

Obscured links are an important threat to the validity of link analysis data. If one or more university sites in our data set used them extensively enough to prevent it being effectively indexed, then it may not be possible to conduct an effective analysis of the set. When using data from a crawler, we have to accept its inability to find many types of automatically generated pages as an unavoidable limitation.

Another concern with web crawler coverage is highlighted by the 'bow-tie' model of the web (Figure 2.1) which treats the web as a mathematical graph and splits it into five named components defined by their connectedness (Broder et al., 2000). This model also applies to individual web sites as well as the whole web. A web crawler starting at a university home page would only be able to reach pages that are directly linked to from the home page or indirectly linked to by repeatedly following links in pages within the site, starting at the home page. In the Broder et al., (2000) terminology, this would normally be the SCC (or largest Strongly Connected Component) and OUT (pages that are accessible from the SCC but do not link back to it). Another explanation for the change in trends observed in this research could thus be that the other components of individual university web sites (i.e., IN, TENDRILS and DISCONNECTED) may have grown, but this trend has not been detected by the crawler. This could be possible for example, if universities were exercising more control over their sites and not allowing as many links from ‘official’ university pages to ‘unofficial’ pages such as personal home pages or clubs, which may have links to the university site, but would not then be linked to by it.

8.6 Conclusions

The main finding of this research is that both the average and median number of static pages for the academic webs of the UK, New Zealand and Australian universities appears to have stabilised, in absolute terms, over a five-year period, even allowing for an expected increase in new web technologies which may have adversely affected the accuracy of the data collection method. This suggests that these academic webs have stopped expanding, and may be tending towards equilibrium. The analysis of individual universities showed that the trend is an average however, and not one that applies to the majority of individual universities. It is difficult to speculate about other countries, although a natural hypothesis now would be to suggest that the number of static pages in their university web sites is likely to have stabilised in countries that were not late adopters of the internet, probably including North America, Western Europe, Taiwan, South Korea and Japan. In addition, other countries may be expected to continue growing exponentially until a natural limit is reached, following the common s-curve or logistic growth model.

Another interesting finding is that the movement of the average number of static pages for UK and Australian webs display striking similarities. This is all the more remarkable given their geographical differences.

These findings have importance to both past and future academic web link analyses in that, if the academic web is indeed stabilising over time, the low rate of variation in academic web site size would lengthen the shelf-life of many webometric results, increasing the time before these results became obsolete. Nevertheless, researchers should be particularly careful with interpretations of international comparisons between sizes, where one nation’s university web sites are in an exponential growth phase whereas the other’s has stabilised.

9 Longitudinal Trends in Academic Web Links

9.1 Introduction

The web is a highly dynamic medium and, while this has obvious advantages in terms of the currency of its content, it is a cause of major concern to web researchers. Due to its continuously evolving nature, the results of any web study may be out of date by the time they reach publication. In the absence of any longitudinal study, web researchers cannot report the results of any trends identified in their web studies as being definitely conclusive, but only as an estimate at any given point in time. This chapter aims to fill this gap by investigating, in some detail, changes in link counts over time for three academic web spaces.

This continues the longitudinal study of the academic web spaces of New Zealand, Australia and United Kingdom universities using data collected as part of an ongoing academic web link database project (Thelwall, 2002/3). While this project has been collecting university link data since 2000, in the context of web analysis this is a long-term perspective and has already been used to provide significant insight into the patterns and relationships inherent in academic hyperlinks (Smith & Thelwall, 2002; Li et al., 2003; Thelwall & Harries, 2004a).

While much research has been carried out on academic web links, and longitudinal studies have been undertaken on Internet web sites and domains (Koehler 1999c; Cho & Garcia-Molina, 2003; Ortega, Aguillo & Prieto, 2006), the research questions in this chapter have been chosen in an attempt to fill a critical gap in current webometrics research. By undertaking a longitudinal study of academic web spaces, it is hoped that patterns and trends in inlinks and outlinks over time, particularly with regard to academic research, can be identified and explained.

9.2 Research Questions

The purpose of this chapter is to identify and track the most significant changes and causes of changes in the academic web space over time, using a case study of three countries for which relevant historical data is available (Australia, New Zealand and the UK). This chapter attempts to gain insights into the stability of results for webometric studies and deals specifically with trends in inlinks and outlinks to and from university web sites in an attempt to answer the following four research questions:

How has the relationship between UK university inlinks and research activity indicators varied over time?

Which universities in each of the three academic web spaces have experienced the greatest increase / decrease in inlinks over the last six years and why?

Can inlink counts be used to assess UK university research activity?

For each academic web space, what percentage of university outlinks change from year to year?

9.3 Methods

The data used during this study takes the shape of text files, one for each university, containing a list of source pages and target hyperlinks. The text files were obtained as part of the ongoing University of Wolverhampton Academic Web Link Database Project (Thelwall, 2002/3) with a collection of national universities' text files forming a database. This paper uses data for the universities of the UK, Australia and New Zealand over a six year period beginning in July 2000. The university text files were then processed using a suite of bespoke programs designed to work with the structure of the text files produced by the crawler, sorting, counting and analysing the link data.

Although the crawls were not taken at exactly the same time period each year attempts were made, especially from 2001 onwards, to crawl each national academic web space every year, and at roughly the same time each year. Given the similarities between each academic web space, the slight time discrepancy seems unlikely to significantly affect the results. Also, the fact that a complete data set of all UK, Australian and New Zealand University links is available precludes the use of random samples, as used in other longitudinal studies (e.g., Koehler, 1999b; Fetterly et al., 2003).

9.4 Results

9.4.1 Correlation with Research Activity

Logarithmic graphs were produced showing linear regression models for UK university site inlinks against research activity (measured as the number of full time faculty members for each university multiplied by that university's RAE rating) for each year between 2000 and 2005. The measure of research activity used here has been used in many previous academic web studies, revealing statistically significant correlations (Smith & Thelwall, 2002; Li et al., 2003; Thelwall, 2002d).

The average RAE rating of the universities was taken from the Times Higher Education Supplement (Mayfield University Consultants, 2001), which averages the grades awarded to each university by the government RAE. This is a peer review, subject-based process that is used to direct Government research funding. Staff numbers were taken from Noble Publishing Higher Education Financial Yearbook (Noble Publishing Co., 1999). It is recognised that the use of static staff number and RAE data is not ideal, especially for a longitudinal study, although staff numbers and RAE averages are relatively stable for most universities and so should not significantly impact upon the results.

The following five logarithmic graphs show linear regression models for UK university site inlinks against research activity (measured as the number of full time faculty members for each university multiplied by that university's RAE rating) for the six year time frame. The raw data used to produce Figures 9.1 – 9.13, and Tables 9.1 and 9.2, can be found at Appendices 2 and 23. Much research has already been carried out using this measure of research activity, revealing statistically significant correlations (Smith & Thelwall, 2002; Thelwall, 2002c; Li et al., 2003). As in chapter 8, the data points representing the London Metropolitan University have been identified as outliers, and have been omitted.

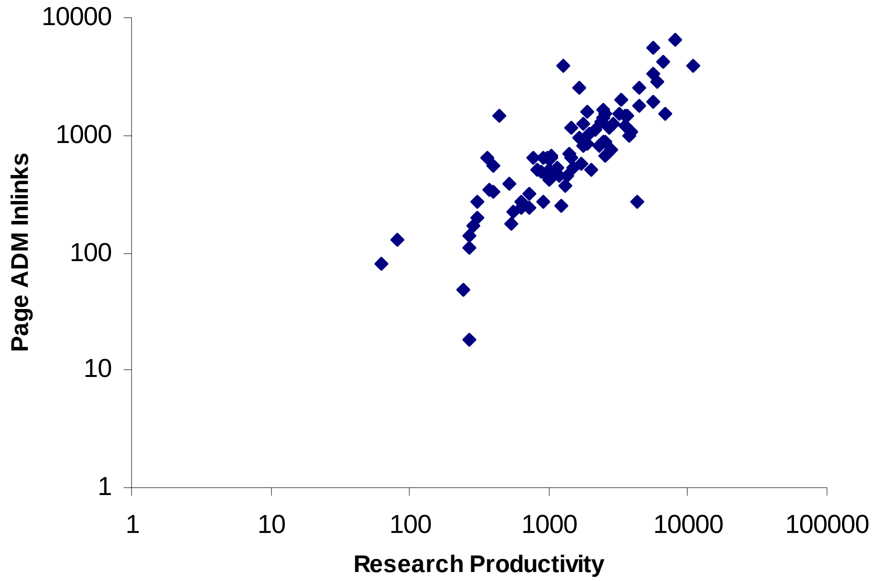

Figure 9.1 UK Page ADM Inlinks against Research Activity for Year 2000

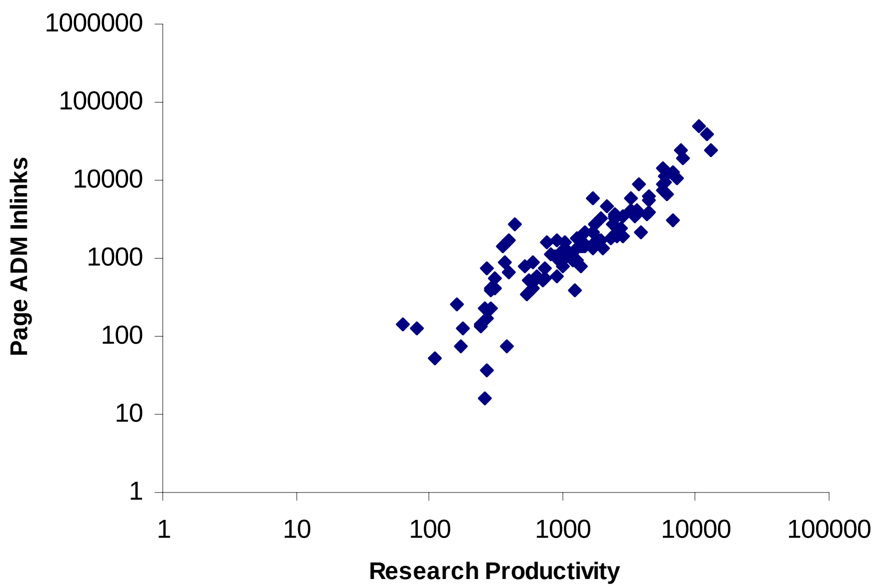

Figure 9.2 UK Page ADM Inlinks against Research Activity for Year 2001

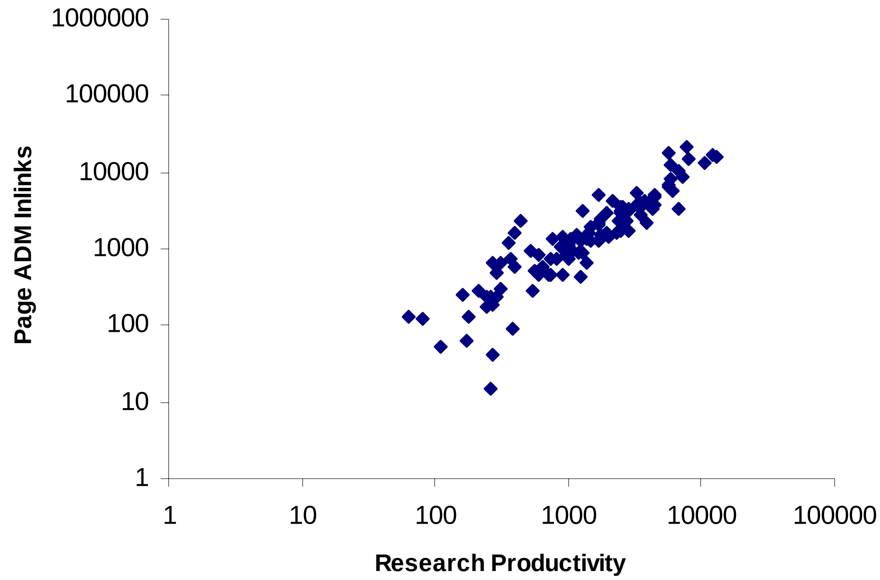

Figure 9.3 UK Page ADM Inlinks against Research Activity for Year 2002

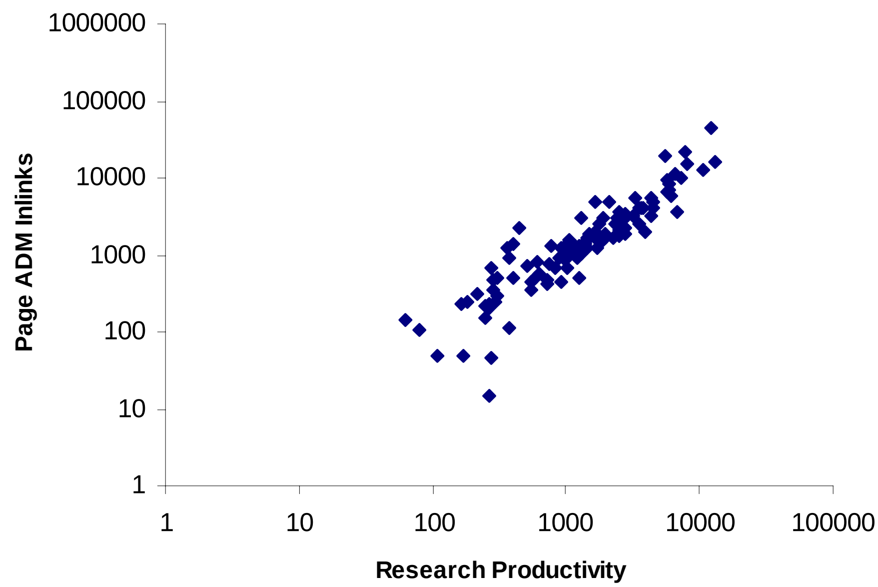

Figure 9.4 UK Page ADM Inlinks against Research Activity for Year 2003

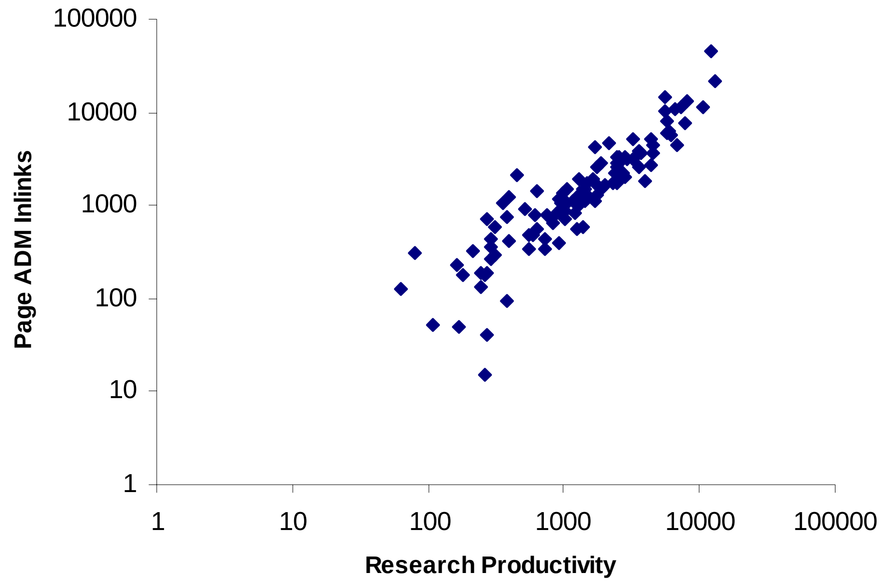

Figure 9.5 UK Page ADM Inlinks against Research Activity for Year 2004

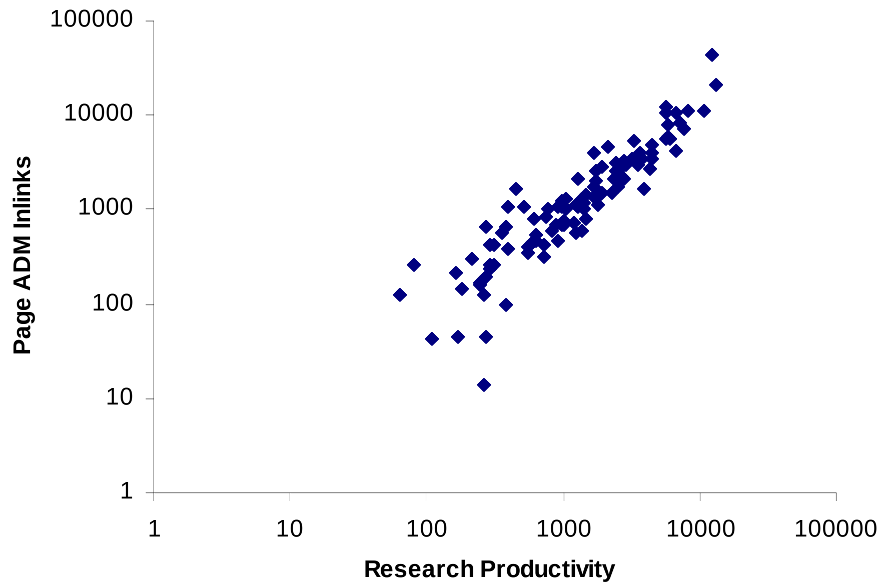

Figure 9.6 UK Page ADM Inlinks against Research Activity for Year 2005

Table 9.1 Spearman's Correlation Coefficient for non-normalised UK logarithmic graphs

Year	Spearman's correlation coefficient, ρ
2000	0.828
2001	0.926
2002	0.925
2003	0.938
2004	0.936
2005	0.944

All graphs displayed correlation coefficients in a very narrow range between 0.828 and 0.944. Levels of correlation greater than 0.7 can be described as high and so, using Spearman's correlation coefficient ρ, it is clear that these graphs all display high levels of correlation.

The following logarithmic graphs, Figures 9.7 – 9.12, show the results for UK university site inlinks divided by the number of full-time academic staff against research activity divided by the number of full-time academic staff. The reason for comparing two indicators divided by faculty numbers is to ensure that both are normalised for size. Bigger universities could be expected to conduct more research and attract more inlinks and so a correlation between total research activity and total inlinks could be explained through both being related to university size. After normalising for size however, another explanation must be sought for any significant correlation found.

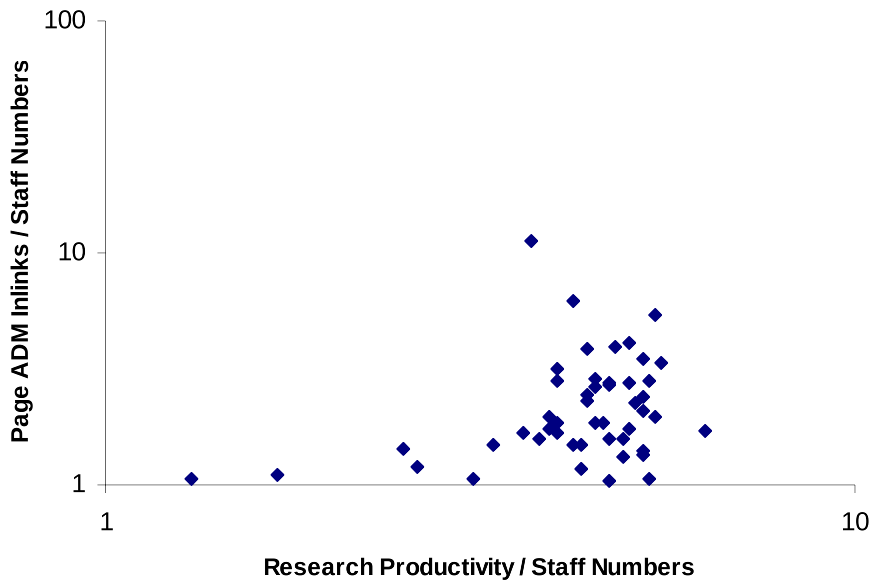

Figure 9.7 UK Page ADM Inlinks divided by Staff Numbers against Research Activity divided by Staff Numbers for Year 2000

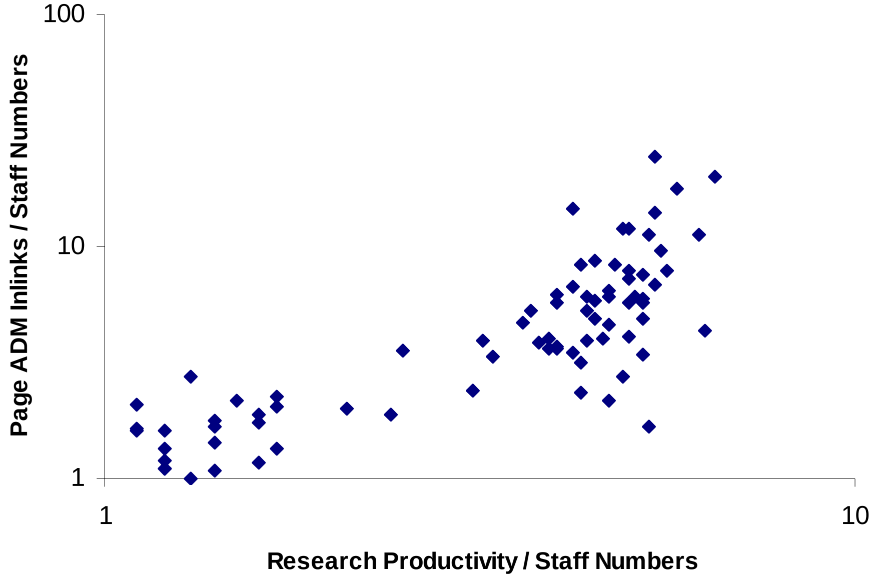

Figure 9.8 UK Page ADM Inlinks divided by Staff Numbers against Research Activity divided by Staff Numbers for Year 2001

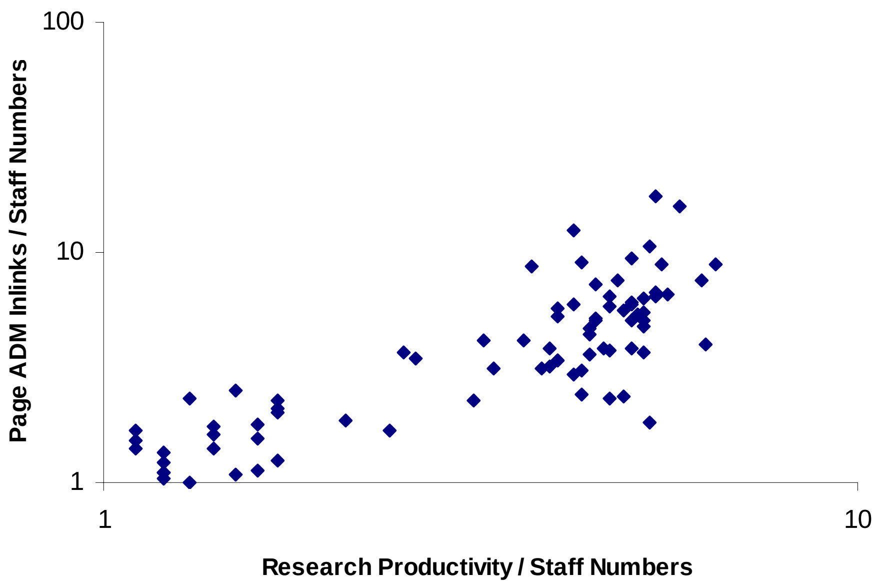

Figure 9.9 UK Page ADM Inlinks divided by Staff Numbers against Research Activity divided by Staff Numbers for Year 2002

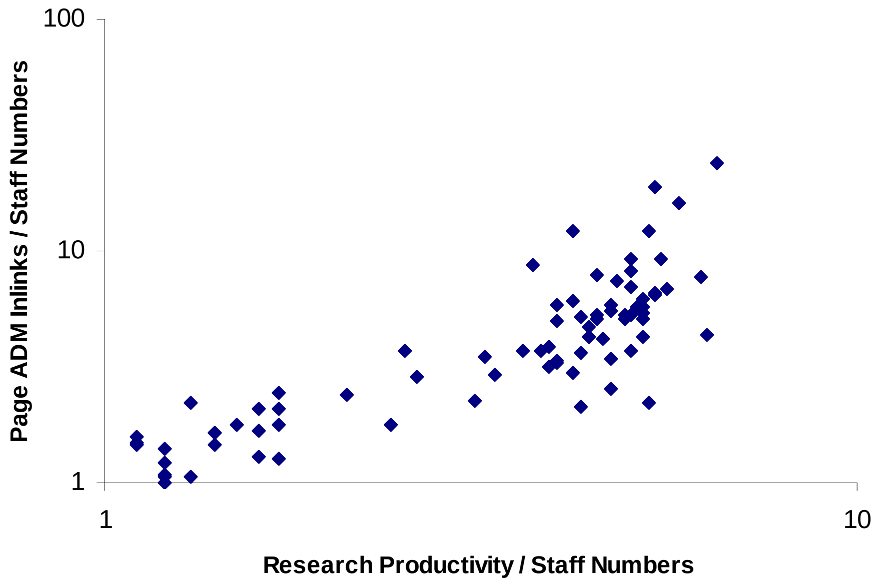

Figure 9.10 UK Page ADM Inlinks divided by Staff Numbers against Research Activity divided by Staff Numbers for Year 2003

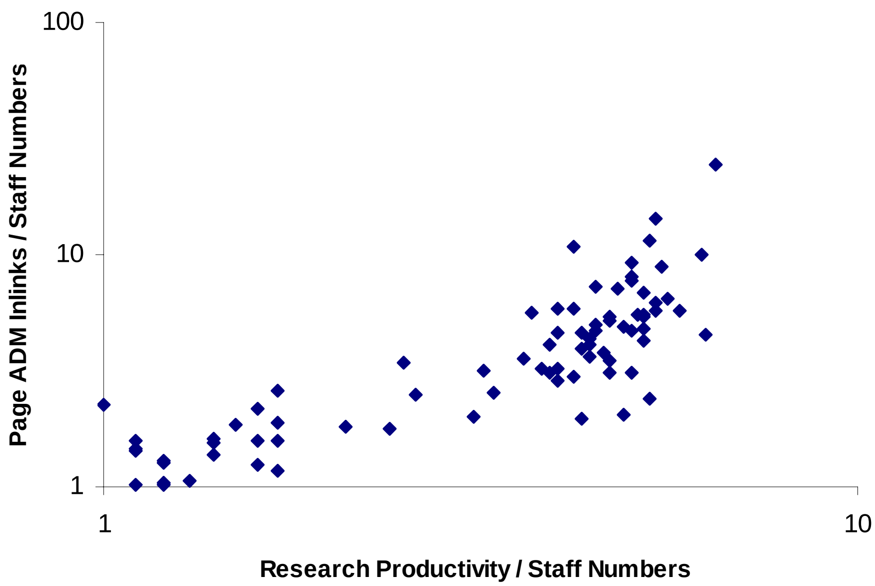

Figure 9.11 UK Page ADM Inlinks divided by Staff Numbers against Research Activity divided by Staff Numbers for Year 2004

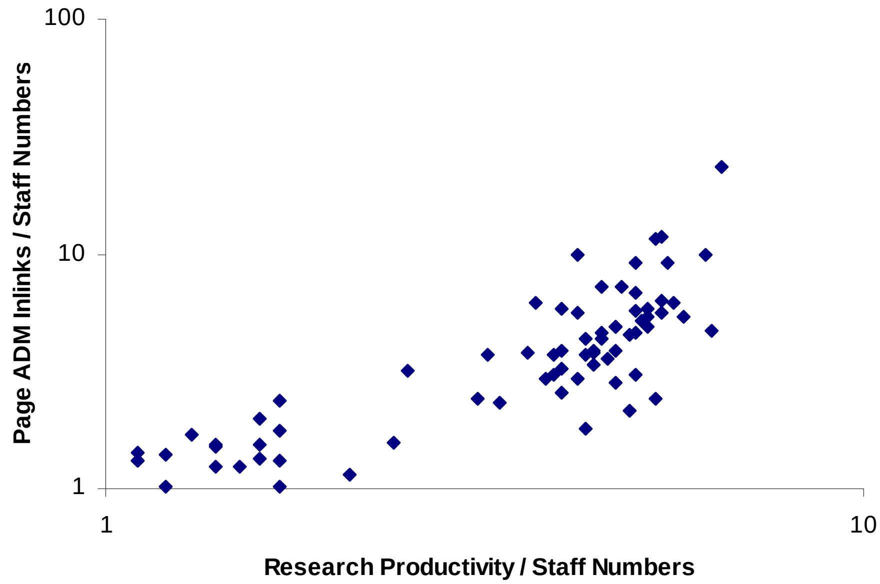

Figure 9.12 UK Page ADM Inlinks divided by Staff Numbers against Research Activity divided by Staff Numbers for Year 2005

Table 9.2 Spearman's Correlation Coefficient for normalised UK logarithmic graphs

Year	Spearman's correlation coefficient, ρ
2000	0.731
2001	0.827
2002	0.832
2003	0.850
2004	0.837
2005	0.845

A summary of the results from the logarithmic graphs is shown below in Figure 9.13. All graphs displayed correlation coefficients in a very narrow range between 0.731 and 0.85. Although the graphs post-normalisation exhibited a lower level of correlation when compared to the non-normalised graphs, they still display high levels of correlation overall.

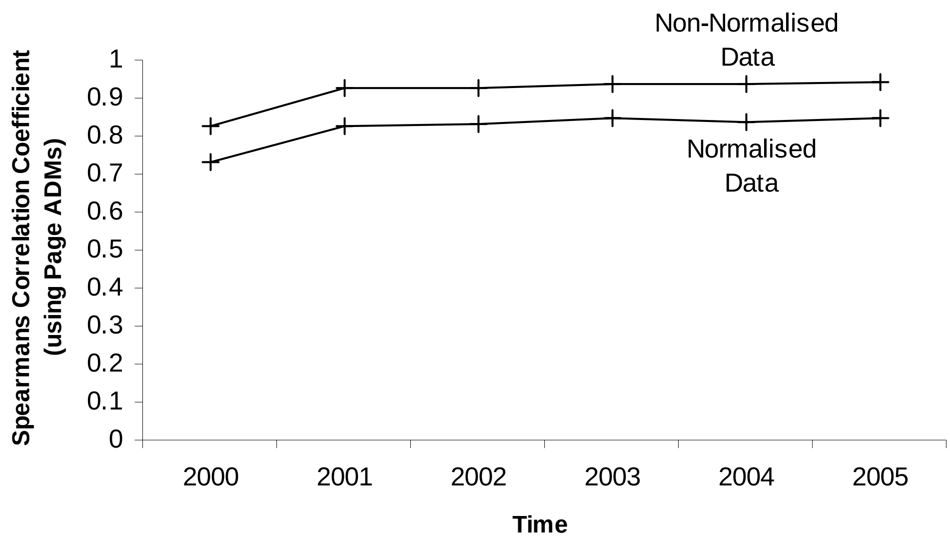

Figure 9.13 Spearman Correlations (using Page ADMs) between Research Activity and Inlink Counts for UK Universities for Normalised and Non-Normalised Data against Time

The same process was undertaken using the Directory ADM (using the raw data given in Appendices 2 and 24) as opposed to the Page ADM with the following results.

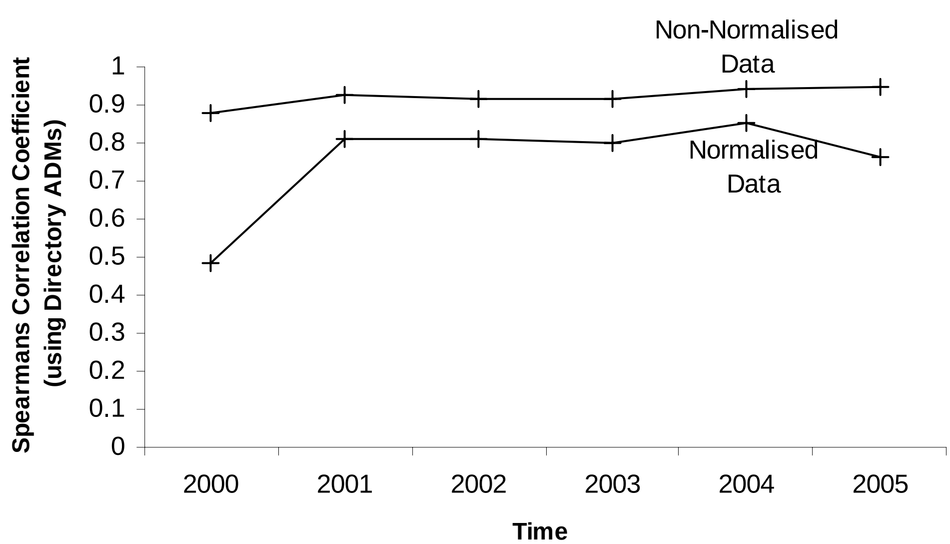

Figure 9.14 Spearman Correlations (using Directory ADMs) between Research Activity and Inlink Counts for UK Universities for Normalised and Non-Normalised Data against Time

While the results above seem to suggest a level of stability for university site inlinks when measured against research activity, it was hypothesised that the number of inlinks would vary considerably for individual universities.

9.4.2 Major Changes in Links

While the results so far have concentrated solely on UK university inlinks (as RAE ratings and staff number data is freely available for the UK), the following tables show the universities which have experienced the greatest percentage increase and decrease in inlinks over the six year period for the UK, Australian and New Zealand academic web spaces, together with the time period in which the change took place. Tables 9.3 and 9.4 below are again produced using the data given in Appendices 21 – 23.

Table 9.3 Percentage Increase of Site Inlinks –
Top Three New Zealand, Australian and UK Universities

University Name	Time Period	Percentage Increase
New Zealand		
Auckland University of Technology	Jul 2000 – Feb 2002	940
Lincoln University	Jul 2000 – Feb 2002	422
Otago University	Jul 2000 – Feb 2002	378
Australia		
Victoria University	Aug 2000 – Jan 2002	3541
University of Melbourne	Aug 2000 – Jan 2002	298
Cowan University	Aug 2000 – Jan 2002	284
UK		
Cardiff University	Jul 2000 – Jul 2001	1196
University College London	Jul 2000 – Jul 2001	1157
University of Reading	Jul 2000 – Jul 2001	807

Table 9.4 Percentage Decrease of Site Inlinks –
Top Three New Zealand, Australian and UK Universities

University Name	Time Period	Percentage Decrease
New Zealand		
Lincoln University	Jan 2003 – Dec 2003	32
Auckland University	Dec 2003 – Jan 2005	16
Waikato University	Dec 2003 – Jan 2005	15
Australia		
Victoria University	Mar 2003 – Feb 2004	93
University of Melbourne	Mar 2003 – Feb 2004	60
Cowan University	Jan 2002 – Mar 2003	55
UK		
University College London	Jul 2001 – Jul 2002	73
Imperial College, University of London	Jun 2003 – Jun 2004	64
Goldsmiths College, University of London	Jun 2003 – Jun 2004	60

The universities showing the greatest percentage increase in inlinks are shown in Table 9.3 while the universities showing the greatest percentage decrease in inlinks are shown in Table 9.4. While some of the reasons for the large increases and decreases in inlinks can be determined by visual examination of the hyperlink text files themselves, the Internet Archive was used extensively to find and display historical university pages in an attempt to better identify the motivation for creating (and deleting) inter-university links, and these are discussed below.

Perhaps the most obvious pattern to emerge from examination of these tables is that, for all three academic web spaces, the universities which underwent the greatest increase in inlinks experienced it during the first time period of the study. This time period, beginning July 2000, shows a phase during which many universities realised the potential and benefits associated with a well designed and functional web site, and therefore enhanced their web presence. This subsequently led to an increase in the number of inlinks pointing to their web sites as most of these links come from other national universities which were also expanding at the time. In New Zealand, Auckland University of Technology, Lincoln University and Otago University experienced the largest increase in inlinks, while in the UK Cardiff University, UCL and Reading University topped the table.

The tables also highlight a remarkable pattern in the Australian academic web space in that the three universities showing the greatest percentage increase in the first time period are the same universities (in the same order) with the greatest percentage decrease over the next two time periods. The 3541% increase in inlinks to

Victoria University up to Jan 2002 is due to other Australian universities, especially the University of Adelaide, linking to the rgmia.vu.edu.au (Research Group in Mathematical Inequalities and Applications) and the sci.vu.edu.au (School of Computer Science and Mathematics) domains. Between Mar 2003 and Feb 2004, the links from the University of Adelaide fell from 5410 to 2 as the university web site underwent reorganisation, and this contributed to a 93% decrease in inlinks to Victoria University. The noted increase for the University of Melbourne over the same time period is due to linking to the unimelb.edu.au/pwebstats/pwebstats.html page (a Perl Web Stats Generator). In Mar 2003 James Cook University and Swinburne University of Technology had a combined total of 13401 links to this page but the software was withdrawn and, by Feb 2004, this had fallen to just 312. The observed increase in inlinks to Cowan University is mainly due to a significant increase in links from the Charles Stuart University to accountancy pages within cowan.edu.au. During the period Jan 2002 to Mar 2003, 34 of the 38 Australian universities experienced a net decrease in the number of inlinks as most national universities consolidated their web sites and Cowan University led the way with a 55% decrease.

In the case of the New Zealand academic web space, the 32% decrease in site inlinks to Lincoln University is largely due to the University of Canterbury removing its links to the www.lincoln.ac.nz/emd directory once the Environmental Management and Design department became unavailable. The 16% decrease experienced by Auckland university can be mainly attributed to the Victoria University of Wellington no longer linking to the www.auckland.ac.nz/lbr/nzp/nzlit2/authors.htm page (a selective list of New Zealand and Pacific authors' works) although this page continues to be available. The 15% decrease experienced by Waikato University can be attributed to a decrease in the number of Victoria University of Wellington pages linking to the www.waikato.ac.nz/library/resources/subject_portal directory. Both of these decreases can be explained by the fact that the Victoria University of Wellington web site underwent a major restructuring during the Dec 2003 to Jan 2005 period, with the number of pages within its site falling from 79241 to 36047.

For the UK universities, the 73% decrease experienced by the University College London can be explained by the fall in links from the University of Warwick to the University College London CATH Protein Structure Classification Database (from 33228 inlinks in Jul 2001 to 50 in Jul 2002). This is due to the database being updated and moved to a different server with a different domain name. Links from the University of Brighton to Imperial College, University of London fell from 13220 in Jun 2003 to 15 in Jun 2004, and this was the major contributory factor in the 64% decrease in inlinks noted. The majority of these links were to foldoc.doc.ic.ac.uk, a free online dictionary of computing which has since moved to http://foldoc.org. This is still affiliated to the Imperial College, University of London, but would not be recognised as such by the web crawler for technical reasons, i.e. as 'ic.ac.uk' no longer forms part of its URL. The 60% decrease in inlinks to Goldsmiths College, University of London (from 806 in Jun 2003 to 2 in Jun 2004) can be traced to a single student with a large number of pages on the City University, London web server for students personal pages, all repeating numerous links to Goldsmiths College, University of London pages.

The results of Tables 9.3 and 9.4 are surprising as, although previous research has claimed that counts of outlinks to universities may vary significantly over short periods of time, the same appears to be true for university inlinks.

9.4.3 Inlinks and Research Activity

Examination of the reasons for significant changes in link behaviour between universities shown in Table 9.3 shows that common reasons for an increase in inlink counts include:

- An increase in web presence i.e. an increase in the number of pages in a university site
- Links to freely available online resources (e.g. databases, programs, dictionaries)

While reasons for the decrease in inlinks shown in Table 9.4 appear to be mainly technical, and include:

- Web site re-organisation (including the introduction of dynamically generating link technology)
- Changes in domain names
- Withdrawal, or movement, of online resources
- Links to personal (non-academic) pages

Therefore, while Figure 9.1 suggests a level of stability over time for UK university site inlinks when measured against research activity, analysing individual UK universities which have experienced significant change from year to year shows that the reasons for these changes are due primarily to web site reorganisation and the introduction (or withdrawal) of online resources such as databases or dictionaries. The fact that these are web-related, rather than research-related factors, would suggest that web links should not be used as a reliable indicator of academic research potential.

9.4.4 Individual Changes in Links

In an attempt to answer the question of what percentage of links change from year to year, the emphasis of this study changes from inlinks to outlinks. To calculate how many outlinks were changed i.e. added or deleted (including instances where the URL was modified), a program was written to count the number of distinct outlinks in a university text file, and then identify the number of duplicate outlinks in subsequent text files. The format of the raw data necessitates the use of outlinks but it could be argued that by running comparison checks to find percentage change between subsequent text files, the overall set of outlinks would be the same as the overall set of inlinks for each academic web space. The results are shown in Table 9.5 below.

Table 9.5 Changes in Outlinks between Years
(expressed as a percentage of the total links in each year)

Year (n)	2002/2003	2003/2004	2004/2005	2005/2006	Average
New Zealand					
Year *n* outlinks also present in year *n*-1. (same)	69%	70%	55%	69%	66%
Year *n* outlinks *not* present in year *n*-1. (new)	31%	30%	45%	31%	34%
Year *n*-1 outlinks not present in year *n*. (missing)	286%	31%	26%	40%	96%
Australia					
Year *n* outlinks also present in year *n*-1. (same)	66%	67%	71%	67%	68%
Year *n* outlinks *not* present in year *n*-1. (new)	34%	33%	29%	33%	32%
Year *n*-1 outlinks not present in year *n*. (missing)	52%	30%	119%	57%	65%
UK					
Year *n* outlinks also present in year *n*-1. (same)	66%	78%	68%	70%	70%
Year *n* outlinks *not* present in year n-1. (new)	34%	22%	32%	30%	30%
Year *n*-1 outlinks not present in year *n*. (missing)	78%	105%	41%	97%	80%

The results for changes in the New Zealand academic web space during the first time period proved to be inconclusive; with data only available in August 2000 for three universities the results were too widely skewed to be considered reliable. Therefore, only data from July 2001 was used for each academic web space and so Table 9.5 shows the percentage change in outlinks for the three academic web spaces for the last four time periods only. It shows the percentage of links which were the same as the previous year (i.e. the percentage of the current year's outlinks which were also present in the previous year) and the percentage of links which were new for this year (i.e. the percentage of the current year's outlinks which were not present in the previous year). These two percentages total 100%. Also shown is the percentage of outlinks which are missing from the previous year (as a percentage of the number of outlinks in the current year to maintain consistency with the other results).

Not shown in Table 9.5, but worthy of mention, are cumulative figures over the four year period. In 2005/2006, the New Zealand academic web space had 24% of the same outlinks it had in 2002/2003 (i.e., 76% new outlinks). Figures for the Australian and UK academic web spaces over the same period showed 31% the same (69% new) and 33% the same (67% new) respectively. This shows that some outlinks are remarkably persistent over time and is consistent with outlinks being gradually, but not systematically, renewed or replaced. The raw data used to produce Table 9.5 can be found in Appendices 25, 27 and 29 while the cumulative data can be found at Appendices 26, 28 and 30.

We can see that the percentage change for outlinks which were the same as the previous year (and hence the percentage change for the number of new links) for all three academic web spaces is within a relatively narrow band, with New Zealand having a 15% spread (between 70% and 55%), the UK a 12% spread (between 78% and 66%) and Australia a 5% spread (between 71% and 66%). To summarise Table 9.5 in general terms we could claim that in most years about two thirds of outlinks are inherited from the previous year and one third are new. In addition, a variable percentage of the previous year's outlinks are lost, with occasional large losses.

Perhaps most significant is the fact that, over the five year period, the average percentage change for outlinks which were the same as the previous year for all three academic web spaces is between 66% and 70%. Consequently, the average percentage change for new outlinks for all three academic web spaces over the five-year period is between 34% and 30%. This is remarkable as each university web site has developed independently, with no formalised guidelines for academic web site development or organisation and with obvious geographical disparity.

9.4.5 Inter-University Domain Links

While other research in this chapter has concentrated on either inlinks or outlinks to and from individual universities, Tables 9.6 – 9.9 below use both to show the top five most popular inter-university links between UK University domains from 2001 – 2004. This information was derived using SocSciBot tools and the data used to construct these tables can be found in Appendix 31.

Figures 9.15 – 9.18 display this information graphically using Pajek diagrams, with the thickness of the links and arrows depicting the number of links and the direction of linking.

Table 9.6 Top Five UK Inter-University Domain Links Jul 2001

From	To	Number of Links
University of Cambridge	University of Oxford	724
University of Oxford	University of Cambridge	559
University of Cambridge	Imperial College, University of London	283
University College London	University of Cambridge	279
University of Glasgow	University of Edinburgh	274

Table 9.7 Top Five UK Inter-University Domain Links Jul 2002

From	To	Number of Links
University of Cambridge	University of Oxford	615
University of Oxford	University of Cambridge	457
University of Edinburgh	University of Cambridge	275
University of Cambridge	Imperial College, University of London	241
University of Glasgow	University of Edinburgh	237

Table 9.8 Top Five UK Inter-University Domain Links Jun 2003

From	To	Number of Links
University of Cambridge	University of Oxford	701
University of Oxford	University of Cambridge	442
University of Glasgow	University of Edinburgh	282
University of Cambridge	University of Edinburgh	280
University of Edinburgh	University of Cambridge	278

Table 9.9 Top Five UK Inter-University Domain Links Jun 2004

From	To	Number of Links
University of Cambridge	University of Oxford	642
University of Oxford	University of Cambridge	445
University College London	University of Cambridge	266
University of Cambridge	University of Edinburgh	260
University of Edinburgh	University of Cambridge	250

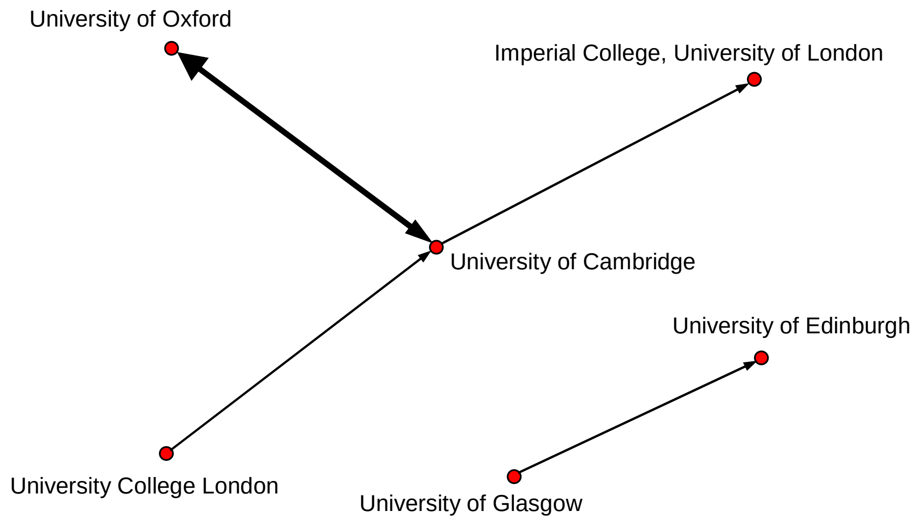

Figure 9.15 Pajek Diagram for Top Five UK Inter-University Domain Links 2001

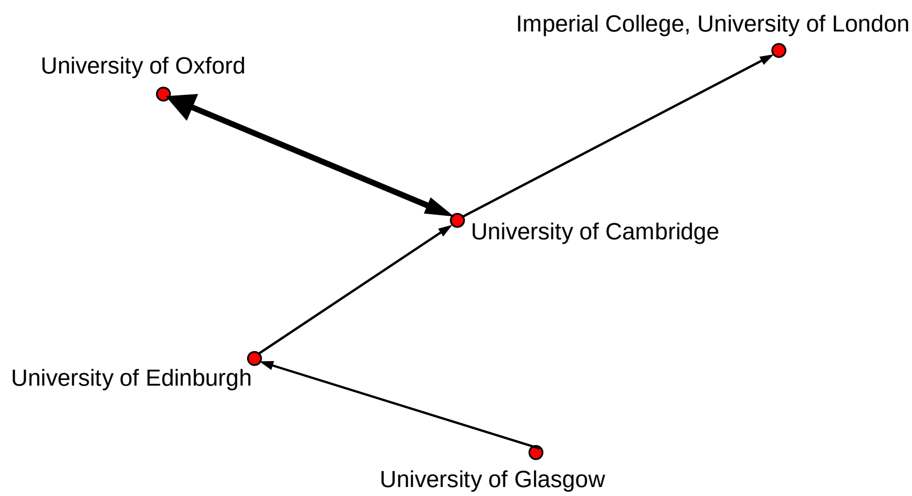

Figure 9.16 Pajek Diagram for Top Five UK Inter-University Domain Links 2002

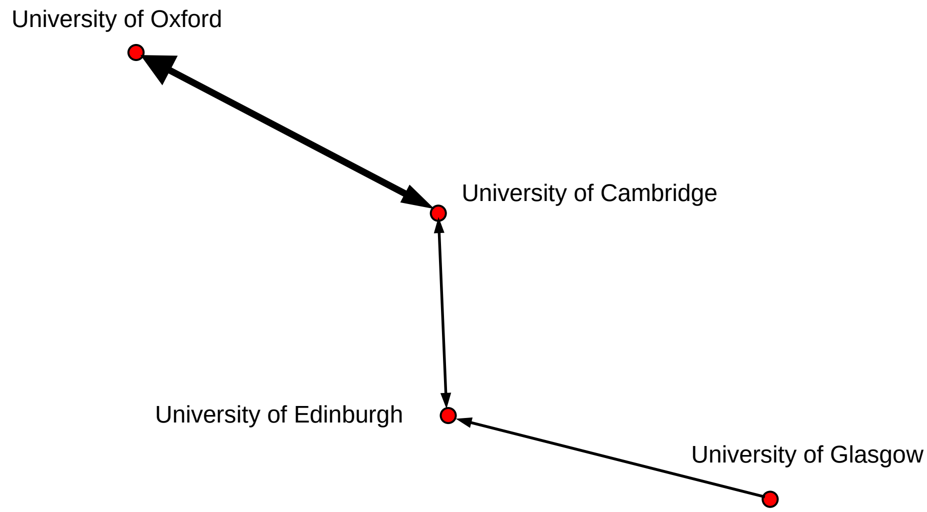

Figure 9.17 Pajek Diagram for Top Five UK Inter-University Domain Links 2003

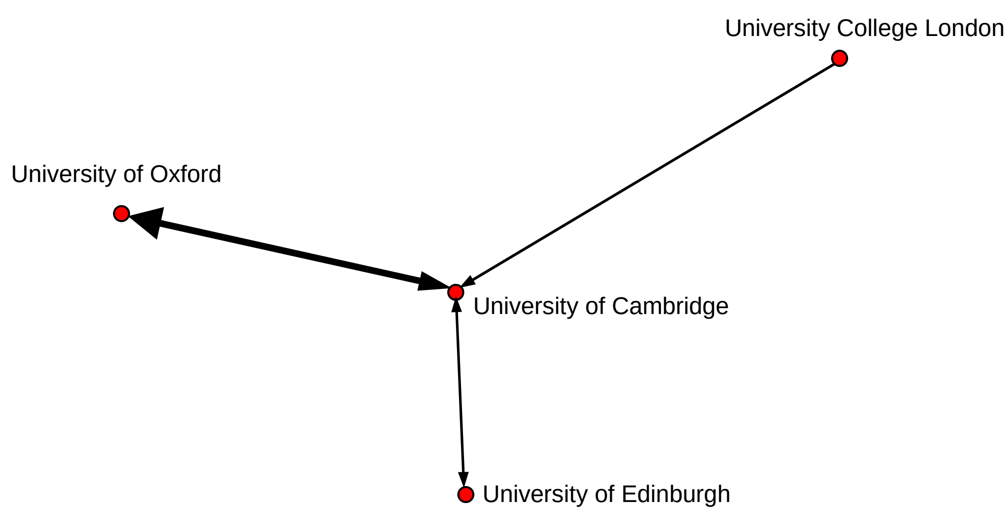

Figure 9.18 Pajek Diagram for Top Five UK Inter-University Domain Links 2004

It is immediately apparent from an examination of Tables 9.6 – 9.9 and Figures 9.15 – 9.18 that links between the 'red-brick' universities dominate these results, with links from the University of Cambridge to the University of Oxford consistently at the top of the tables, and with links from the University of Oxford to the University of Cambridge consistently in second place. It is interesting that, apart from always appearing in the top two places, the University of Oxford does not feature in the remaining places of any of the tables.

Other observations include the regular occurrences of links from the University of Glasgow to the University of Edinburgh, echoing the findings of Thelwall (2002f) that there is a correlation between geographic location and research-related interlinking between universities, with particular emphasis on the Scottish establishments. There also appears to be evidence of increasingly important interlinking between the Universities of Cambridge and Edinburgh, in both directions.

The Pajek diagrams in Figures 9.15 – 9.18 make it easy to see that the University of Cambridge has held an important, central role in UK inter-university interlinking throughout the period of this study. It is suggested that the number of domain links between these universities could be evidence of strong research collaborative links although further research would be required to prove this. We can see however that all of the universities in these tables have high RAE ratings. In any future metrics-based UK research rating exercise, it may be worth considering a PageRank-type approach in which the rank of the university is not determined by the number of links alone, but also by the relative ranking of the institutions linking to it.

9.5 Discussion

The results for non-normalised and normalised data shown in Figure 9.13 show that the correlation between the average number of site inlinks and research activity has remained relatively constant over the time period in question. A significant correlation between link count statistics and another independent measure is evidence that there is some pattern in the link data, and is suggestive of a connection between the two data types. In addition, these results suggest that 'staff number*RAE rating' is a reliable, stable measure of research activity. However, it is important to remember that a statistically significant correlation between two variables does not imply that one is the cause of the other, as there may be unrelated factors that influence both (Vaughan, 2001).

From Tables 9.3 and 9.4 it is immediately apparent that the largest percentage increases all fall within the first time period for all three academic web spaces. This echoes previous findings which show an increase in both average site size and average inlink count for all three academic web spaces over the period Jul 2000 – Feb 2002 (Payne & Thelwall, 2007a). This would appear to be a period when UK, Australian and New Zealand universities were enhancing their web presence. They appear to have increased the size of their respective web sites, adding more pages and consequently more links to other national universities. Indeed, every university in all three academic web spaces saw its inlink count increase during this period and, following this period of growth, all three academic web spaces then entered a period of stabilisation and consolidation.

While no formalised attempt to categorise links or pages of the type carried out in Wilkinson et al., (2003) and Bar-Ilan (2004c) is made for this study, a cursory inspection of the data suggests that the majority of the universities experiencing a significant increase in inlinks appear to gain work-related (including research-related) links, as opposed to links of a purely social, recreational or superficial nature. Other studies warn of the dangers associated with considering links between universities as equivalent to citations, as only about 1% of inter-university links target content equivalent to that of a journal article, although around 90% seem to link to pages with some academic nature, as opposed to purely administrative or recreational pages (Wilkinson et al., 2003). This seems consistent with this study as a significant number of the causes identified for major increases and decreases in inlinks can be attributed to popular shared online university resources. In only one instance, Goldsmiths College, University of London, can a change in the number of inlinks be shown to be entirely social in nature. Perhaps academic sites provide an opportunity to discover patterns within the specific depth of the links (Vasileiadou & Van Den Besselaar, 2006), e.g. the hierarchy of a department could be reflected in the depth of links (we would find the personal homepages of scientists working in that department at a deeper level).

Another interesting point to note is that of the 132 UK, Australian and New Zealand universities for which data was available each year from 2000 to 2005, all have experienced a net increase in the number of inlinks over this period. The greatest increase was experienced by the Auckland University of Technology, rising from 5 inlinks in Jul 2000 to 83 in Jan 05. Again, this 1560% increase is exaggerated by the small number of New Zealand universities and consequently the small number of links between them, but this should not detract from the fact that every university during the course of this study has experienced an overall escalation in the number of inlinks to their site.

From the results shown in Figure 9.1, it appears that the correlation between UK universities' research activity and inlink counts has remained relatively constant from year to year. However, analysing the reasons for significant changes in individual UK universities from Tables 9.3 and 9.4 shows that the majority of changes can be attributed to technical, web-based factors rather than research-related factors. This would seem to suggest that, from a longitudinal perspective, web links should not be used as a reliable indicator of academic research potential.

Shifting the emphasis of the study from inlinks to outlinks in Table 9.5 shows that the overall percentage change of outlinks in the year of study which also appeared in the previous year for the academic web spaces of New Zealand, Australia and the UK was 66%, 68% and 70% respectively. This is a significant finding as it could be stated that on average 66% – 70% of outlinks in all three academic web spaces do not change year on year. Brewington and Cybenko (2000), after downloading around 100,000 pages per day between March and November 1999, found that for pages downloaded six times or more, 56% did not change at all over the duration of the study, while 4% changed every single time. Although these are comparable results, this study concentrated on changes in web page content, not hyperlinks. Also shown in Table 9.5 is the percentage of links in the year of study lost from the previous year but there does not appear to be an obvious pattern within these statistics, with results varying widely for all three national academic web spaces.

It would be interesting to ascertain whether any of these results would be significantly affected by the introduction of alternative link analysis methods. ADMs (Thelwall, 2002e) may have the potential to produce better web link analysis results. While the current study has concentrated on identifying trends in inlinks and outlinks at the university level for each of the three academic web spaces, the adoption of directory or domain ADMs may have added more stability by reducing the impact of some of the significant changes in link counts experienced by individual universities.

9.5.1 Limitations

There are a number of limitations to this study inherent in the design of the data collection method. Web crawlers operate by following links and are limited in that they can only find pages that they are allowed to visit, are linked to or already know about and are linked to in a way in which the crawler can extract from the linking page. The number of pages found will depend upon the site, the crawler design and the parameters under which the crawler is operating. A significant limitation in the information science web crawler used to collect the data used in this study is its inability to crawl dynamically generated URLs, as many universities have, over the period of this study, adopted technologies which integrate them into their web sites.

This chapter deals with both inlinks and outlinks. It is sometimes argued that inlinks are more useful as indicators than outlinks, because outlinks are under the control of the site owners whereas inlinks are not. An additional technical problem with site outlink counts is that they depend upon a single site crawl, and are therefore more liable to crawler coverage problems than inlink counts, which are totalled from a number of different crawls. For example, if one site is not covered well by a crawler because an important area of the site has pages in a format that cannot be crawled, then this will have a big impact upon the outlink count for that site, but only a small impact on the inlink count of all other sites, which will lose the inlinks that were missed from the badly crawled site.

Another limitation is due to the fact that only static staff number and RAE data were available for the duration of this study. While it is recognised that this is not ideal, especially for a longitudinal study, staff numbers and RAE averages are relatively stable for most universities and so should not significantly impact upon the results.

9.6 Conclusions

This research focuses on inlink and outlink count variations over time in academic web spaces. It is important to know as much as possible about the changes the web, and web links, experience over time because the rate of variation impacts upon the shelf-life of webometric results.

In terms of individual outlinks, in the case of the three national academic web spaces in this study, it seems that about two thirds (66% to 70%) of outlinks remained the same from year to year for all three academic web spaces, although this apparent stability conceals large individual differences, such as a high percentage of individual outlinks disappearing from one year to the next (Table 9.5).

When counts of outlinks from academic web sites are compared over time then the changes observed could be expected to include large jumps for individual universities. Big increases can easily occur if a large collection of pages is added and big decreases can occur when sets of old pages are deleted. Previous research has assumed that inlinks to the same sites should be steadier, at least in relative size, as they depend upon the content of pages from a range of other sites. An exception to this occurs when one university introduces or removes a collection of pages with a disproportionately large collection of external links, and this phenomenon is apparent in this study by the variance in links due to the introduction or removal of popular large online resources such as portals and databases. Anomalies are therefore to be expected in any comparison of sources of link counts over time even if there is an otherwise linear trend.

The results presented here support this with evidence of relative stability for university site inlinks, as measured against research activity over time, but there are surprisingly large fluctuations in these inlinks at the individual university level. The majority of the causes for these changes are due to web-based, not research-based, factors and hence this supports, from a longitudinal perspective, previous assertions that inlink counts alone should not be used as a reliable indicator of academic research potential for individual universities, although they can be effective at identifying general trends. The results also suggest that comparisons between different inlink counts for the same set of academic web sites are unreliable, even if there is only a short time period between the data collection dates. In particular, if comparing two similar webometric papers produced within a year of each other then it would still not be safe to assume that their raw data was similar. This has far-reaching implications for the replicability and comparability of webometrics research, which undermines the potential of the field to compare techniques and reach an agreement on the best ones.

10 Discussion

10.1 Introduction

While each individual chapter can be viewed as a stand-alone academic study in its own right, with its own specific results, discussions and conclusions, the aim of this chapter is to summarise the main findings of the thesis and to identify original, recurring trends and patterns within this body of research.

10.2 Significant Contributions

Chapter 5, in answering the research question '*Is the relationship between academic web and research activity indicators best modelled by a linear trend or a power law?'*, validates previous assumptions and lays the foundation for future research, including that carried out in chapter 9 with regards to the relationship between academic site size and research activity indicators. Although not specifically a longitudinal study, the results from chapter 5 hold importance as a step towards understanding the phenomenon of academic web linking and developing metrics to extract useful information. The fact that the relationship between academic web site size and research activity is best modelled by a linear trend and not a power law validates the results of previous research into academic web interlinking which has assumed that web data should follow a linear trend (Thelwall, 2002e; Thelwall, 2002d; Thelwall & Wilkinson, 2003b). It also justifies the design and use of web metrics such as the WIF which rely on linear data, as Katz (2000) has shown that specially modified 'scale-independent' indicators would need to be developed for data that obeys a power law. However, many of the 'rich gets richer' type of power law behaviour observed within the web has concentrated on the distribution of overall web links over time, and is not strictly applicable to the collection of academic text file hyperlink structures used within the current study.

While previous studies have shown that more accurate results can be obtained through the use of the domain or directory ADM (Thelwall, 2002d; Thelwall and Wilkinson, 2003b; Payne and Thelwall, 2004; Thelwall, 2004b; Thelwall and Harries, 2004a), chapter 6 shows that these results are consistent over time, and across different countries academic web spaces. Although correlation between the number of pages in a site, and the number of inlinks / outlinks to and from a university site is not a new finding, chapter 6 shows that, while strong correlations do indeed exist between pages and links across all three academic web spaces over the last six years, consistently higher results are obtained when the web links are aggregated at the directory and especially domain level and, in doing so, answers the research question '*Which Alternative Document Models give the most consistent results when applied to the UK, Australian and New Zealand academic web spaces 2000 – 2006?'*. However, the success of the domain model is not conclusive enough to be able to claim that it is the sole definitive model for link analysis research, although the results do show that it is significantly better than the default page model. Conversely, aggregating at the site (or university) level appears to conclusively provide less reliable results than using the page as the standard unit of measure, and this finding holds true over all three academic webs and for each time period analysed over the last six years. These results can be partially explained by the fact that the use of ADMs will balance anomaly elimination with loss of data. A higher level of aggregation will eliminate link anomalies, but will also lose data due to this aggregation. For this study, the directory and domain models have emerged as the most suitable for producing high levels of correlation, as it may be that the page ADM does not sufficiently address the anomalies while the site ADM, due to the high level of aggregation, loses too much data. In any case, these positive results for the directory and domain ADM appear to strengthen the case for using web link analysis as a tool with the potential to reveal underlying trends in academic website interlinking.

Chapter 7 addresses the research question '*How and why does the distribution of types of academic web links change over time?*' and considers whether university web sites publish the same kind of information and use the same kind of hyperlinks year on year. This is an important issue from the perspective of interpreting the results of academic link analyses, because changes over time in link types would force interpretations of link analyses to also change over time. This chapter is a first attempt to combine a link classification exercise with a longitudinal study, finding that there are significant differences in the way the distribution of certain types of academic web links change over time. Significant increases in 'research oriented', 'social/leisure' and 'superficial' links were identified as well as notable decreases in 'technical' and 'personal' links. Some of these differences can be explained by general changes in the management of university web sites i.e. 'research-oriented' and 'social/leisure' links and some by more wide-spread Internet trends such as dynamic pages, blogs and social networking i.e. 'personal' and 'superficial'. The decline in 'technical' links over the years is an unexpected finding, although this may be explained by the amalgamation of previously separate online resources. The increase in the proportion of research-oriented links discovered in chapter 7 is particularly hopeful for future link analysis research. The analogy between journal citations and hyperlinks is at the very heart of many webometric studies and so, if this increase continues, there may eventually be enough to have a significant overall influence on university hyperlinking although at the moment, the numbers are very low.

Chapter 8 attempts to answer the research question '*What is the trend over time for the average web site size, and average inlink count, of UK, Australian and New Zealand academic webs?*'. Among the many longitudinal-based results presented here, the most significant finding is that both the average and median number of static pages for the academic webs of the UK, New Zealand and Australia appears to have stabilised, in absolute terms, over a five-year period even allowing for an expected increase in new web technologies which may have adversely affected the accuracy of the data collection method. This suggests that these academic webs have stopped expanding, and may be tending towards equilibrium. The analysis of individual universities shows that this trend is an average however, and not one that applies to the majority of individual universities. It is difficult to speculate about other countries, although a natural hypothesis would be to suggest that the number of static pages in their university web sites is likely to have stabilised in countries that were not late adopters of the Internet, probably including North America, Western Europe, Taiwan, South Korea and Japan. In addition, other countries may be expected to continue growing exponentially until a natural limit is reached, following the common s-curve or logistic growth model. Another interesting finding is that the movement of the average number of static pages for UK and Australian webs display striking similarities and this is all the more remarkable given their geographical differences. The findings from chapter 8 have importance to both past and future academic web link analyses in that, if the academic web is indeed stabilising over time, the low rate of variation in academic web site size would lengthen the shelf-life of many webometric results, increasing the time before these results became obsolete.

Chapter 9 focuses on outlink and inlink count variations over time in academic web spaces and answers the following four research questions:

'How has the relationship between UK university inlinks and research activity varied over time?'

'Which universities in each of the three academic web spaces have experienced the greatest increase / decrease in inlinks over the last six years and why?'

'Can inlink counts be used to assess UK university research activity?'

'For each academic web space, what percentage of university outlinks change from year to year?'

The results provide evidence of relative stability for UK university site inlinks as measured against research activity over time, but there are surprisingly large fluctuations in these inlinks at the individual university level. The majority of the causes for these changes are due to web-based, not research-based, factors and hence this supports, from a longitudinal perspective, previous assertions that inlink counts alone should not be used as a reliable indicator of academic research potential for individual universities, although they can be effective at identifying general trends. The results also suggest that comparisons between different inlink counts for the same set of academic web sites are unreliable, even if there is only a short time period between the data collection dates. In particular, if comparing two similar webometric papers produced within a year of each other then it would still not be safe to assume that their raw data was similar. This has far-reaching implications for the replicability and comparability of webometrics research, which undermines the potential of the field to compare techniques and reach an agreement on the best ones. In terms of individual outlinks, it seems that about two thirds (66% to 70%) of outlinks remain the same from year to year for all three academic web spaces, although this apparent stability conceals large individual differences, such as a high percentage of individual outlinks disappearing from one year to the next.

10.3 Recurring Trends

Other studies aside, this body of research has identified several common, recurring trends during the compilation of this thesis.

Firstly, distinct similarities in the characteristics of both the UK and Australian academic webs have been identified. Chapter 6 shows that for these two countries, all four ADMs exhibit similar correlations with site size using both inlinks and outlinks, and that the strength of these correlations appears to be increasing over time. Chapter 8 found that many of the graphs produced for UK and Australia universities exhibited striking similarities, both in their shape and behaviour. The similarities between these two webs are all the more remarkable given that there exists no formalised recommendations for web page creation or guidance for official academic web sites in any individual country, let alone in two countries with such an obvious geographical disparity.

The academic web of New Zealand is seen to exhibit markedly different characteristics, both in terms of ADM correlation (chapter 6) and graphs of average site size over time (chapter 8). These differences may be explained by the fact that there is a much reduced sample size for New Zealand, i.e. there are only 8 New Zealand universities in the Academic Web Link Database, compared to around 110 for the UK and 38 for Australia. This makes it much easier for a New Zealand university with a large number of web site pages, such as the Victoria University of Wellington, to dominate the results. However, this may have an alternative explanation as Smith and Thelwall (2002) compared linking between UK, Australian and New Zealand universities and found that New Zealand was relatively isolated on the web, suggesting that the academic web in New Zealand is somewhat insular; relatively well interconnected but less well known internationally. This finding is also in line with the results of a previous bibliometric study for journals (Glänzel, 2001) which shows New Zealand to be one of the most isolated of the advanced nations in terms of international scientific co-authorship.

It was also expected that the link classification exercise carried out in chapter 7 would highlight similar inter-country differences. However, examination of the results shows no identifiable disparity between the link types of the UK, New Zealand and Australian academic webs. In this case, all three webs appear to display the same characteristics, even when viewed from a longitudinal perspective.

Chapter 5, in answering the important question of whether the relationship between academic web size and interlinking data is best modelled by a linear trend or a power law, demonstrated high levels of correlation between research activity indicators and a measure of UK university site size for 2002. Chapter 9 confirms these results and introduced a longitudinal perspective, showing that the correlation between the average number of site inlinks and research activity has remained relatively constant between 2000 and 2006. A significant correlation between link count statistics and another independent measure is evidence that there is some pattern in the link data, and is suggestive of a connection between the two data types. In addition, these results suggest that ‘staff number*RAE rating’ is a reliable, stable measure of research activity.

The decrease in the number in ‘technical’ links between the years 2000 and 2006 shown in chapter 7 was a surprising result. One would expect that the number of links to online libraries, databases, journals and applications to have increased over time but this does not appear to be the case. One possible explanation, borne out by the findings of chapter 8 is that the number of actual technical links has remained the same but, because the number of source pages of universities within the three academic webs experienced a marked increase between 2000 and 2001, as a percentage of overall links, 'technical' links may appear to have decreased. Additionally, reasons for the decrease in inlinks shown in Table 9.4 appear to be mainly technical, and include the withdrawal, or movement, of 'technical' online resources.

While no formalised attempt to categorise links or pages was made during chapter 9, a cursory inspection of the data suggests that the majority of the universities experiencing a significant increase in inlinks appear to gain work-related (including research-related) links and this supports the findings of chapter 7 where significant increases in 'research oriented', 'social/leisure' and 'superficial' links were identified. Also, in more general terms, the increase in dynamic pages identified in Table 8.7 may provide an explanation for the changes in link types identified during chapter 7 as it is possible that certain types of links historically found in static pages are now more often found in dynamic pages and therefore concealed from the majority of web crawlers. These findings highlight an important issue for Internet longitudinal studies in that the introduction of new technologies such as social networking and dynamic pages must be accounted for if valid comparisons of web indicators are to be made.

10.4 Research Originality

Philips (1992) attempted to list ways in which originality in PhD work can be shown and this thesis attempts to conform to the following 7 definitions:

Carrying out empirical work which hasn't been done before. The current study uses mainly empirical methods and, while the main results from chapters 5 and 6 may be regarded as validations of previous assumptions, the findings presented in chapters 7, 8 and 9 are innovative and original.

Using already known material but with a new interpretation. Although correlation between the number of pages in a site, and the number of inlinks / outlinks to and from this site is not a new finding, chapter 6 shows that, while strong correlations do indeed exist between pages and links, consistently higher results are obtained when the web links are aggregated at the directory and domain level across all three academic web spaces over the last six years.

Trying out something in this country which has previously only been done in other countries. Chapter 7 addresses concerns expressed in Bar-Ilan (2004b) over only considering links between the academic institutions of one country (Israel) by considering the links within the three academic web spaces of the UK, New Zealand and Australia.

Taking a particular technique and applying it to a new area. The use of ADMs as a means to reduce anomalies is a well established practice within webometrics. However, their use in a longitudinal context, as in chapter 6, is a novel application of the technique.

Looking at areas that people in the discipline haven't looked at before. While current webometric research concentrates on static web pages, chapter 8 also includes a study of dynamic pages to show that the average and median number of pages for the academic webs of the UK, New Zealand and Australian universities appears to have stabilised over a five-year period even allowing for an expected increase in new web technologies.

Adding to knowledge in a way that hasn't been done before. While there have been many studies involving academic webs spaces, and much work has been carried out on the web from a longitudinal perspective, the main body of the current study concentrates on filling a critical gap in current webometric research by combining the two and undertaking a longitudinal study of academic webs.

Being cross-disciplinary and using different methodologies. Chapter 7 presents a first attempt to address the important issue of how different types of academic web links change over time, blending web-based longitudinal data (e.g. Koehler 1999b; 2002; 2004, Fetterly et al., 2003) with a link classification exercise (e.g. Bar-Ilan 2004b; 2004c, Wilkinson et al., 2003).

11 Conclusions and Future Work

11.1 Conclusions

In his PhD thesis, Björneborn (2004) criticised the review chapter 'Measuring the Internet' (Molyneux & Williams, 1999) because, due to the rapid changes on the Internet, much of its coverage was unavoidably out of date. Also, in realising that the data set used in his study constituted of a frozen snapshot of the publicly available university link structure data, he specifically called for future longitudinal studies of academic web sites to be undertaken. Similarly, Kitchens and Mosley (2000) questioned the value of printed Internet references due to the ephemeral nature of the web while Spink et al., (2001) commented that many Internet studies are old, if not ancient, by the time they are published. This body of research shows that, because the web is changing every day, researchers should not report their results as being correct and definitive about the nature of the web, but only as being estimates at a given point in time. In other words, in the absence of any longitudinal study, web research can give insights into the way the web is used but cannot give any long-term conclusions because of the dynamic nature of the web. The above studies all demonstrate the importance of web-based longitudinal research, as any study attempting to identify the amount by which the web changes over time would be of value to other web researchers, if only to give them an indication of how long their results are likely to remain current for.

This thesis fills the critical gap in current webometric research highlighted above by undertaking a longitudinal study of academic web spaces. The main body of this research concentrates on applying established webometric and statistical methods in a longitudinal framework and is a first step towards identifying and explaining how the academic hyperlink structure changes over time. It explores the changing nature of linking between universities in order to gain a better understanding of academic web spaces.

Chapter 5 validates the assumptions of previous webometric research and allows future researchers to carry out their studies with the prior knowledge that the relationship between academic web and research activity indicators is best modelled by a linear relationship. This relationship is also shown to be relatively steady over time (chapter 9), although there are surprisingly large fluctuations in inlinks at the individual university level. As the majority of the causes for these changes are due to web-based (as opposed to research-based) factors this shows that, from a longitudinal perspective, inlink counts alone should not be used as a reliable indicator of academic research potential for individual universities although, as in this case, they can be effective at identifying general trends.

The results from chapter 6 bring together assertions from previous studies and demonstrate conclusively that domain and directory ADMs successfully reduce the impact of anomalies in web data and, while other studies have already shown that more accurate results can be obtained through their use, it is shown here that these results are consistent over time, and across different countries' academic web spaces. Webometricians should be aware of the fact that aggregation at the domain or directory level, as opposed to the page or site level, consistently produce more reliable results over time and that the use of the directory or domain ADM should be encouraged where possible.

The increase in the proportion of citation-equivalent research-oriented links discovered in chapter 7 is particularly hopeful for future link analysis research. The often invoked analogy between journal citations and hyperlinks is at the very heart of many webometric studies and, if this trend continues, there may eventually be enough to have a significant overall influence on university hyperlinking. This finding suggests that webometrics should become increasingly effective as a tool to track research online.

The results in chapter 9 also suggest that comparisons between different inlink counts for the same set of academic web sites are unreliable, even if there is only a short time period between the data collection dates. In particular, if comparing two similar webometric papers produced within a year of each other then it would still not be safe to assume that their raw data was similar. This has far-reaching implications for the replicability and comparability of webometrics research, which undermines the potential of the field to compare techniques and reach an agreement on the best ones.

A significant finding of this research is that, in terms of size and linking patterns, the academic web spaces of the UK, Australia and New Zealand have been relatively stable since 2001, (even allowing for an expected increase in new web technologies), although this does hide a constant renewing of old pages and links. Chapter 8 suggests that these academic webs have stopped expanding and may be tending towards equilibrium. Although it is difficult to speculate about other countries, a natural hypothesis would be to suggest that the number of static pages in the university web sites of countries that were early adopters of the Internet is likely to have stabilised, while in other countries it may be expected to continue growing until a natural limit is reached. These findings have importance to both past and future academic web link analyses in that, if the academic web is indeed stabilising over time, the low rate of variation in academic web site size would lengthen the shelf-life of many webometric results, increasing the time it takes for these results to become obsolete.

It may have been previously assumed that webometric results would have a fairly short useful shelf-life, due to the dynamic and fast moving nature of the environment in which the research is undertaken. However, this study has shown that webometric results are likely to have a surprisingly long shelf-life, perhaps closer to five years than one year. However, this would be dependant upon the nature of the of webometric study undertaken. For example, a study of the relative sizes of the university systems in two or more developed countries over a five year period should be reliable as the total numbers should be relatively stable. But, as chapter 9 shows, the results for individual universities are not stable, even from one year to the next and so, although the size trends are constant, the results for individual universities are highly unstable and should not be relied upon. Similarly, this study shows that the links themselves are altering, although over a five year period a substantial proportion of these links will remain the same. Nevertheless, chapter 7 shows that the type of links is slowly changing and that later results may connect more closely to research than earlier results, although the change is not large enough to be significant, even over a five year period. Overall then, for most webometric purposes (with the exception of monitoring individual universities), five years seems a reasonable life span for webometric results. However, this would be reliant on no modern web technologies becoming widely adopted which may render the current results obsolete. It is also worth re-iterating that these results are based entirely on developed, English-speaking nations and that further research would be needed to clarify whether the results presented here are transferable.

While there have already been many studies involving academic webs spaces, and much work has been carried out on the web from a longitudinal perspective, this thesis concentrates on filling a critical gap in current webometric research by combining the two and undertaking a longitudinal study of academic webs. In comparison with previous web-related longitudinal studies this thesis makes a number of novel contributions. Some of these stem from extending established webometric results, either by introducing a longitudinal aspect (looking at how various academic web metrics such as research activity indicators, site size or inlinks change over time) or by their application to other countries. Other contributions are made by combining traditional webometric methods (e.g. combining topical link classification exercises with longitudinal study) or by identifying and examining new areas for research (for example, dynamic pages and non-HTML documents).

The importance of the inherent properties of web links means that the type of research undertaken in this thesis should have a promising future. In addition, the results so far show that robust, durable information can be extracted from large scale comparisons of web links between academic institutions and this strengthens the case for the use of webometrics as a tool to reveal underlying trends in academic interlinking. Conducting a study of this type may also help webometricians define and understand the factors which influence the stability and reliability of their research and thus, in addition to determining the shelf-life of webometric research, it may be possible to increase it by identifying and optimising the factors which influence it.

11.2 Future Work

The most obvious avenues for future research following the current longitudinal study would be either to extend the time period of the study or to extend its scope to include academic web spaces for other countries, especially developing nations or non-English speaking developed nations. It is possible that the domain structure in other countries would be different, and so the results would not necessarily extend. Future research confirming the results for other countries would strengthen the findings.

It would also be interesting to see whether similar results could be obtained from a longitudinal study of different web spaces, either other academic webs such as secondary schools, or non-academic areas of the web such as commercial or government webs. A problem with both would be the unavailability of a relevant, comprehensive data set, although this kind of research may be possible in the future using the Internet Archive.

More specifically, several promising directions for future research have emerged naturally during the course of this study.

11.2.1 Longitudinal Motivation Studies into Hyperlink Creation

In addition to chapter 7 of the current study, previous research has emphasised the variety in link creation motivations (Wilkinson et al., 2003; Bar-Ilan, 2004b, 2004c; Harries et al., 2004) and has highlighted the importance of longitudinal studies into the motivation behind hyperlink creation. This variety makes link classification and motivation studies difficult, but longitudinal studies would be vital in order to determine how types of links change over time and to develop a future understanding of how link counts should be interpreted.

11.2.2 Wider Social Sciences Research

Since the web is not exclusively an academic space, it can be used in wider social science research both as an object in its own right (e.g. to study online communities) and as an easily accessible source of information about offline phenomena that happen to be reflected in the web. The dramatic increase in the popularity of so-called 'Web 2.0' sites such as Flickr, YouTube and MySpace (which is suggested as a primary reason for some of the changes in link types identified in chapter 7) is characterised by user-driven content and offers new opportunities to apply traditional bibliometrics approaches to emerging technologies. Although this technology is relatively new, a link-based longitudinal study would not only provide information on how traditional hyperlinks are being affected by this phenomenon over time, but may also provide some insight into the changing nature of social interaction over the Internet introduced by these new, social-networking web sites.

11.2.3 Blog Link Analyses

Web logs (blogs) are online diaries maintained by millions of web users around the world and, as such, are an enormous repository of information. There is extensive linking within and between blogs and therefore they seem to be a particularly promising medium for the extension of link analysis techniques. In fact, blog link counts are already used to compile a daily list of the top 100 most popular blogs (www.blogstreet.com/top100.html). While the lack of quality control over blog posts and the lack of natural topic organisation found in academic journals may be problematic when it comes to extracting useful information from blog link counts, a longitudinal blog link analysis could be useful in general terms by providing information about the development of blogging, (e.g. Cohen & Krishnamurthy, 2006) or more specifically by providing data about the spread of individual topics (e.g. Thelwall, 2006b).

11.2.4 Dynamically Generated Pages

This body of research is one of few webometric studies to address the issue of dynamically generated pages (although Brewington & Cybenko, 2000 attempt to calculate the number of these pages in their study by counting how many pages change on every repeat observation) and Figure 8.7 indicates an (unsurprising) increase in their use over time. Due to the programmatic differences and issues with data collection methods no effort was made to amalgamate the graphs in Figures 8.1 and 8.7, because of the potential to produce misleading results. However, given their recent popularity, it would be interesting to conduct a study dedicated solely to examining the use of dynamic web pages. Webometric data collection methods have historically relied on commercial search engines or dedicated web crawlers, neither of which deals specifically well with the issue of dynamically generated pages. It may be the case that, if webometric studies were adjusted to include data for dynamically generated pages, this would significantly affect their results.

11.2.5 Page Types

The popularity of Internet trends such as blogs and social networking appears to be having an affect on how the distribution of web links changes for individual universities over time and this can be expected to continue.

The decline in 'technical' links over the years is surprising although this may be explained by the amalgamation of previously separate online resources and further study would be needed to more clearly identify the reason for this. The increase in the proportion of research-oriented links discovered in chapter 7 is particularly hopeful for future link analysis research. The analogy between journal citations and hyperlinks is at the very heart of many webometric studies and so if this increase continues, there may eventually be enough to have a significant overall influence on university hyperlinking.

References

ADAMIC, L. A. & HUBERMAN, B. A. (2000) Power-Law Distribution of the World Wide Web. *Science*, 287, pp. 2115a.

ADAMIC, L. A. & HUBERMAN, B. A. (2001) The Web's hidden order. *Communications of the ACM,* **44**(9), pp. 55-59.

ADAMIC, L. A. & HUBERMAN, B. A. (2002) Zipf's Law and the Internet. *Glottometrics* 3, pp. 143-150.

AGUILLO, I. F. (1998). STM information on the Web and the development of new Internet R&D databases and indicators. *Online Information '98: Proceedings*, pp. 239-243.

AGUILLO, I. F. (2002) Cybermetrics : definitions and methods for an emerging discipline. *Séminaires de l'ADEST*, Paris, 14 February, 2002. Available: http://www.upmf-grenoble.fr/adest/seminaires/ISIDRO/Cybermetrics.ppt

AGUILLO, I. F. (2004) Cybermetrics of the Indian Universities. International Conference on Webometrics, Informetrics and Scientometrics and 5th COLLNET Meeting, Roorkee, India.

AJIFERUKE, I. & WOLFRAM, D. (2004a) Informetric Modelling of Internet Search and Browsing Characteristics. *Canadian Journal of Information and Library Science*. **28**(1), pp. 1-16.

AJIFERUKE, I. & WOLFRAM, D. (2004b) Modelling the characteristics of web page outlinks. *Scientometrics*, **59**(1), pp. 43-62.

ALBERT, R. & BARABÁSI, A. (2002) Statistical mechanics of complex networks. *Reviews of Modern Physics*, **74**(1), pp. 47-97.

ALBERT, R., JEONG, H. & BARABÁSI, A. L. (1999) The Diameter of the World-Wide Web. *Nature*, 401, pp. 130.

ALMIND, T.C. (1995) *Informetrisk undersøgelse af den danske del af World WideWeb*. Masters Thesis, Danmarks Biblioteksskole, Copenhagen.

ALMIND, T. C. & INGWERSEN, P. (1997) Informetric Analyses on the World Wide Web: Methodological Approaches to 'Webometrics'. *Journal of Documentation*, **53**(4), pp. 404-426.

AN, L. & QIU, J. P. (2004) Research on the relationships between Chinese journal impact factors and external web link counts and web impact factors. *Journal of Academic Librarianship,* **30**(3), pp. 199-204.

BARABÁSI, A. (2001) The physics of the Web. *Physics World*, July 2001. Available: http://physicsweb.org/article/world/14/7/9

BARABÁSI, A. & ALBERT, R. (1999) Emergence of scaling in random networks. *Science*, **286**, pp. 509-512.

BARABÁSI, A., ALBERT, R. & JEONG, H. (1999) Mean-Field Theory for Scale-Free Random Networks. *Physica*, **272**, pp. 173-187.

BARABÁSI, A., ALBERT, R. & JEONG, H. (2000) Scale-free characteristics of random networks: The topology of the World-Wide Web. *Physica A*, **281**, pp. 69-77.

BAR-ILAN, J. (1998) On the overlap, the precision and estimated recall of search engines: A case study of the query "Erdos". *Scientometrics*, **42**(2), pp. 207-208.

BAR-ILAN, J. (1999) Search Engine Results over Time - A Case Study on Search Engine Stability. *Cybermetrics*, 2/3. Available: http://www.cindoc.csic.es /cybermetrics/articles/v2i1p1.html

BAR-ILAN, J. (2000) Results of an extensive search for "S&T indicators" on the Web--A content analysis. *Scientometrics*, **49**(2), pp. 257-277.

BAR-ILAN, J. (2001) Data Collection Methods on the Web for Informetric Purposes – A Review and Analysis. *Scientometrics*, **50**(1), pp. 7-32.

BAR-ILAN, J. (2002a) How much information do search engines disclose on the links to a web page? A longitudinal case study of the 'cybermetrics' home page. *Journal of Information Science,* **28**(6), pp. 455-466.

BAR-ILAN, J. (2002b) Methods for measuring search engine performance over time. *Journal of the American Society for Information Science and Technology*, **53**(4), pp. 308-319.

BAR-ILAN, J. (2004a) Self-linking and self-linked rates of academic institutions on the web. *Scientometrics*, **59**(1), pp. 29-41.

BAR-ILAN, J. (2004b) What do we know about links and linking? A framework for studying links in academic environments. *Information Processing and Management*, **41**(4), pp. 973-986.

BAR-ILAN, J. (2004c) A microscopic link analysis of academic institutions within a country - the case of Israel. *Scientometrics,* **59**(3), pp. 391-403.

BAR-ILAN, J. (2004d) The use of Web search engines in information science research. *Annual Review of Information Science and Technology*, **38**, pp. 231-288.

BAR-ILAN, J. (2005) Expectations versus reality – Search engine features needed for Web research at mid 2005. *Cybermetrics*, **9**(1), paper 2. Available: http://www.cindoc.csic.es/cybermetrics/articles/v9i1p2.html.

BAR-ILAN, J. (2007) Google Bombing from a Time Perspective. *Journal of Computer-Mediated Communication*, **12**(3), pp. 910-938.

BAR-ILAN, J. & PERITZ, B. C. (2002) Informetric theories and methods for exploring the Internet: an analytical survey of recent research literature. *Library Trends*, **50**(3), pp. 371-392.

BAR-ILAN, J. & PERITZ, B. C. (2004) Evolution, continuity, and disappearance of documents on a specific topic on the web: a longitudinal study of "Informetrics". *Journal of the American Society for Information Science and Technology*, **55**(11), pp. 580-590.

BAR-ILAN, J., LEVENE, M. & MAT-HASSAN, M. (2006) Methods for evaluating dynamic changes in search engine rankings: a case study. *Journal of Documentation*, **62**(6), pp. 708-729.

BAR-ILAN, J., MAT-HASSAN, M. & LEVENE, M. (2006) Methods for comparing rankings of search engine results. *Computer Networks,* **50**(10), pp. 1448-1463.

BARJAK, F. & THELWALL, M. (2006) A statistical analysis of the web presences of European life sciences research teams. Paper presented at the International Conference on Science, Technology and Innovation Indicators. History and New Perspectives, Lugano 15-17 November 2006.

BARJAK, F., LI, X. & THELWALL, M. (2007) Which factors explain the web impact of scientists' personal homepages? *Journal of the American Society for Information Science and Technology,* **58**(2), pp. 200-211.

BAR-YOSSEF, Z., KUMAR, R., BRODER, A. & TOMKINS, A. (2004) Sic Transit Gloria Telae: Towards an Understanding of the Web's Decay. In *Proceedings International WWW Conference*, New York, USA.

BEAULIEU, A. & SIMAKOVA, E. (2006) Textured Connectivity: an ethnographic approach to understanding the timescape of hyperlinks. *Cybermetrics*, **10**(1), paper 5. Available: http://www.cindoc.csic.es/cybermetrics/articles/v10i1p5.html

BENBOW, S. M. P. (1998) File Not Found: the problem of changing URLs for the World Wide Web. *Internet Research*, **8**(3), pp. 247-250.

BERNERS-LEE, T. (1997) Realising the full potential of the Web. Available: http://www.w3.org/1998/02/Potential.html

BERNERS-LEE, T. & CAILLIAU, R. (1990) *WorldWideWeb: Proposal for a hypertext project*. Available: http://www.w3.org/Proposal.html

BHARAT, K., CHANG, B., HENZINGER, M. & RUHL, M. (2001) Who links to whom: Mining linkage between web sites. *Proceedings of ICDM 2001*, pp. 51-58.

BJÖRK, B. & TURK**,** Z. (2000) How scientists retrieve publications: an empirical study of how the Internet is overtaking paper media. *The Journal of Electronic Publishing*, **6**(2). Available: http://www.press.umich.edu/jep/06-02/bjork.html

BJORN-ANDERSEN, N. & ELLIOT, S. (2005) E-Business Perceptions Versus Reality: A Longitudinal Analysis of Corporate Websites. *Lecture Notes In Computer Science*, **3590**, pp. 178-187.

BJÖRNEBORN, L. (2001a) *Shared outlinks in webometric co-linkage analysis: a pilot study of bibliographic couplings on researchers' bookmark lists on the Web,* Royal School of Library and Information Science.

BJÖRNEBORN, L. (2001b) Small-world linkage and co-linkage. *Proceedings of the 12th ACM Conference on Hypertext and Hypermedia.* New York: ACM Press. pp. 133-134.

BJÖRNEBORN, L. (2004) *Small-World Link Structures across an Academic Web Space - a Library and Information Science Approach.* Royal School of Library and Information Science, Copenhagen, Denmark. (PhD Thesis).

BJÖRNEBORN, L. (2006) 'Mini small worlds' of shortest link paths crossing domain boundaries in an academic Web space. *Scientometrics***,** **68**(3), pp. 395-414.

BJÖRNEBORN, L. & INGWERSEN, P. (2001) Perspectives of Webometrics. *Scientometrics,* **50**(1), pp. 65-82.

BJÖRNEBORN, L. & INGWERSEN, P. (2004) Towards a basic framework for Webometrics. *Journal of the American Society for Information Science and Technology,* **55**(14), pp. 1216-1227.

BOLLEN, J., VAN DE SOMPEL, H., SMITH, J. & LUCE, R. (2005) Toward alternative metrics of journal impact: A comparison of download and citation data. *Information Processing and Management*, **41**(6), pp. 1419-1440.

BORGMAN, C. & FURNER, J. (2002) Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology*, **36**, pp. 3-72.

BOSSY, M. J. (1995) The last of the litter: "Netometrics". *Solaris*, 2 ('Les sciences de l'information : bibliométrie, scientométrie, infométrie'). Presses Universitaires de Rennes. Available: http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/d02/2bossy.html

BOTAFOGO, R. A., RIVLIN, E. & SHNEIDERMAN, B. (1992) Structural analysis of hypertexts: identifying hierarchies and useful metrics. *ACM Transactions on Information Systems*, **10**(2), pp. 142-180.

BREWINGTON, B. E. & CYBENKO, G. (2000) How dynamic is the Web? *Computer Networks*, **33**(1–6), pp. 257–276.

BRIN, S. & PAGE, L. (1998) The Anatomy of a Large Scale Hypertextual Web Search Engine. *Computer Networks and ISDN Systems*, **30**, pp. 107-117.

BRODER, A., KUMAR, R., MAGHOUL, F., RAGHAVAN, P., RAJAGOPALAN, S., STATA, R., TOMKINS, A. & WIENER, J. (2000) Graph Structure in the Web. *Computer Networks*, **33**(1-6), pp. 309-320.

CALADO, P., CRISTO, M., GONÇALVES, M. A., S. de MOURA, E. RIBEIRO-NETO, B. & ZIVIANI, N. (2005) Link-based similarity measures for the classification of Web documents. *Journal Of The American Society For Information Science And Technology*, **57**(2), pp. 208-221.

CHAKRABARTI, S., DOM, B., KUMAR, R., RAGHAVAN, P., RAJAGOPALAN, S., TOMKINS, A., GIBSON, D. & KLEINBERG, J. (1999) Mining the link structure of the World Wide Web. *IEEE Computer*, **32**(8), pp. 60-67.

CHAKRABARTI, S., JOSHI, M. M., PUNERA, K. & PENNOCK, D. M. (2002) The Structure of Broad Topics on the Web. *WWW2002 Conference*. Available: http://www2002.org/CDROM/refereed/338

CHAN, D., CHUA, K., LECKIE, C. & PARHAR, A. (2003) Visualisation of Power-Law Network Topologies. *Proceedings of the Eleventh IEEE International Conference on Networks* (ICON 2003), Sydney, Australia, pp. 69-74.

CHEN, C., NEWMAN, J., NEWNAM, R. & RADA, R. (1998) How did University Departments Interweave the Web: A Study of Connectivity and Underlying Factors. *Interacting with computers*, **10**, pp. 353-373.

CHENG, Y. & LIU, N. C. (2006) A first approach to the classification of the top 500 world universities by their disciplinary characteristics using scientometrics. *Scientometrics***, 68**(1), pp. 135-150.

CHI, E. H., PITKOW, J., MACKINLAY, J., PIROLLI, P., GOSSWEILER, R. & CARD, S. K. (1998) Visualizing the evolution of Web ecologies. *Proceedings of Human Factors in Computing Systems, CHI 98.* ACM Press. pp. 400-407.

CHO, J. & GARCIA-MOLINA, H. (2003) Estimating frequency of change. *ACM Trans. Internet Techn,* **3**(3), pp. 256-290.

CHU, H. (2005) Taxonomy of inlinked Web entities: What does it imply for webometric research? *Library & Information Science Research*, **27**(1), pp. 8-27.

CHU, H., HE, S. & THELWALL, M. (2002) Library and information science schools in Canada and USA: A Webometric perspective. *Journal of Education for Library and Information Science*, **43**(2), pp. 110-125.

COHEN, E. & KRISHNAMURTHY, B. (2006) A short walk in the Blogistan. *Computer Networks,* **50**(5), pp. 615-630.

COOK, T.D. & CAMPBELL, D.T. (1979) Quasi-Experimentation: Design and Analysis for Field Settings. Chicago, Illinois: Rand McNally College Publishing.

COOLEY, R., MOBASHER, B. & SRIVASTAVA, J. (1997) Web mining: information and pattern discovery on the World Wide Web. *Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97).* Available: http://citeseer.nj.nec.com/cooley97web.html

COTHEY, V. (2002) A longitudinal study of World Wide Web users' information-searching behaviour. *Journal of the American Society for Information Science and Technology,* **53**(2), pp. 67-78.

COTHEY, V. (2006) What do hyperlinks mean: the value of hyperlink-networks as indicators of knowledge production. *Cybermetrics*, **10**(1), discussion paper 1. Available: http://www.cindoc.csic.es/cybermetrics/articles/v10i1c1.html

COTHEY, V., AGUILLO, I. & ARROYO, N. (2006) Operationalising “Websites”: lexically, semantically or topologically? *Cybermetrics*, **10**(1), paper 3. Available: http://www.cindoc.csic.es/cybermetrics/articles/v10i1p3.html

CRONIN, B. (1984) *The Citation Process. The role and significance of citations in scientific communication*. London: Taylor Graham.

CRONIN, B. (2001) Bibliometrics and Beyond: Some Thoughts on Web Based Citation Analysis. *Journal of Information Science*, **27**(1), pp. 1-7.

CRONIN, B., SNYDER, H. W., ROSENBAUM, H., MARTINSON, A. & CALLAHAN, E. (1998) Invoked on the Web. *Journal of the American Society for Information Science,* **49**(14), pp. 1319-1328.

CROWSTON, K. & WILLIAMS, M. (2000) Reproduced and emergent genres of communication on the World Wide Web. *Information Society*, **6**, pp. 201-215.

CUI, L. (1999) Rating Health Websites using the Principle of Citation Analysis: A Bibliometric Approach. *Journal of Medical Internet Research*, **1**(1), e4. Available: http://www.jmir.org/1999/1/e4/

DING, C., ZHA, H., HE, X., HUSBANDS, P. & SIMON, H. (2002) Link analysis: hubs and authorities on the World Wide Web. LBNL Tech Report 47847. May 7, 2001 Available: http://www.nersc.gov/research/SCG/cding/papers_ps/hits3.ps

DOROGOVTSEV, S. N., MENDES, J. F. F. & SAMUKHIN, A.N. (2000) Structure of growing networks with preferential linking. *Physical Review Letters*, **85**(21), pp. 4633-4636.

DOUGLIS, F., FELDMANN, A., KRISHNAMURTHY, B. & MOGUL, J. (1997) Rate of change and other metrics: A live study of the World Wide Web. *Proceedings of the Symposium on Internet Technologies and Systems*, Monterey, California. Available: http://www.usenix.org/publications/library/proceedings/usits97/full_papers/douglis_rate/douglis_rate.pdf

DOWNIE, J. S. (1996) Informetrics and the World Wide Web: A case study and discussion. *Proceedings of the 24th Annual Conference of the Canadian Association for Information Science*, 2-3 June 1996, Toronto, Ontario. pp. 130-141.

DUMAIS, S. & CHEN, N. (2000) Hierarchical classification of web content. *Proceedings of ACM SIGIR conference.*

EASTERBY- SMITH, M. THORPE, R. & LOWE, A. (1991) *Management Research: An Introduction*. London: Sage Publications.

EGGHE, L. (1998) Mathematical relations between impact factors and average number of citations. *Information Processing and Management*, **24**(5), pp 567-576.

EGGHE, L. (2000) New Informetric Aspects of the Internet: Some Reflections – Many Problems. *Journal of Information Science*, **26**(5), pp. 329-335.

EGGHE, L. (2005) Expansion of the field of informetrics: Origins and consequences. *Information Processing and Management*, **41**(6), pp. 1311-1316.

EGGHE, L. (2006) Expansion of the field of informetrics: The second special issue. *Information Processing and Management*, **42**(6), pp. 1405-1407.

EGGHE, L. & ROUSSEAU, R. (2002) Co-citation, bibliographic coupling and a characterization of lattice citation networks. *Scientometrics*, **55**(3), pp. 349-361.

ELMER, G. (2006) Re-tooling the network: Parsing the Links, Codes, and Commands of the Web World. *Convergence: The International Journal of Research into New Media Technologies,* **12**(1) pp.9-19.

FABA-PÉREZ, C., ZAPICO-ALONSO, F., GUERRERO-BOTE, V. & DE MOYA-ANEGÓN, F. (2005) Comparative analysis of webometric measurements in thematic environments. *Journal Of The American Society For Information Science And Technology*, **56**(8), pp. 779-785.

FALOUTSOS, M., FALOUTSOS, P. & FALOUTSOS, C. (1999) On Power Law Relationships of the Internet Topology. *Computer Communication Review*, **29**, pp. 251-262.

FANG, Y. & ROUSSEAU, R. (2001) Lattices in citation networks: an investigation into the structure of citation graphs. *Scientometrics*, **50**(2), pp. 273-287.

FETTERLY, D., MANASSE, M., NAJORK, M. & WIENER, J. (2003) A large scale study of the evolution of web pages. *Proceedings of the twelfth international conference on World Wide Web*, pp. 669-678.

FLAKE, G. W., LAWRENCE, S., GILES, C. L. & COETZEE, F. M. (2002) Self-organization and identification of web communities. *IEEE Computer*, **35**(3), pp. 66-71.

FOMENKOV, M., KEYS, K., MOORE, D. & CLAFFY, K. (2004) Longitudinal study of Internet traffic in 1998-2003. *Proceedings of the winter international symposium on Information and communication technologies,* **58,** pp. 1-6.

FOOT, K. A., SCHNEIDER, S. M., DOUGHERTY, M., XENOS, M. & LARSEN, E. (2006) Analyzing Linking Practices: Candidate Sites in the 2002 US Electoral Web Sphere. *Journal of Computer-Mediated Communication*, **8**(4). Available: http://jcmc.indiana.edu/vol8/issue4/foot.html

FRANDSEN, T. F. (2005) Journal interaction: A bibliometric analysis of economics journals. *Journal of Documentation*, **61**(3), pp. 385-401.

FRY, J. (2006) Studying the Scholarly web: How disciplinary culture shapes online representations. *Cybermetrics*, **10**(1), paper 2. Available: http://www.cindoc.csic.es/cybermetrics/articles/v10i1p2.html.

FÜRNKRANZ, J. (1998) Using links for classifying web-pages: Technical Report OEFAI-TR-98-29'. Austrian Research Institute for Artificial Intelligence. Available: http://citeseer.nj.nec.com/153148.html

GALITSKY, B & LEVENE, M. (2002) On the economy of Web links: simulating the exchange process. *First Monday,* **9**(1). Available: http://firstmonday.org/issues/ issue9_1/galitsky/index.html

GARFIELD, E. (1955) Citation indexes for science: a new dimension in documentation through association of ideas. *Science*, **122**, pp. 108-111.

GARFIELD, E. (2006) The history and meaning of the Journal Impact Factor. *Journal of the American Medical Association,* **295**(1), pp. 90-93.

GARRIDO, M. & HALAVAIS, A. (2003) Mapping networks of support for the Zapatista Movement: Applying Social Network Analysis to study contemporary social movements. In: M. McCaughey & M. Ayers (Eds.), *Cyberactivism: online activism in theory and practice*, New York: Routledge, pp. 165-184.

GERMAIN, C.A. (2000) URLs: uniform resource locators or unreliable reliable resource locators? *College and Research Libraries*, **61**(4), pp. 359-365.

GIBSON, D., KLEINBERG, J. & RAGHAVAN, P. (1998) Inferring Web Communities From Link Topology. *Hypertext 98: Ninth ACM Conference on Hypertext and Hypermedia*, ACM, New York, USA.

GLÄNZEL, W. (2001) National characteristics in international scientific co-authorship relations. *Scientometrics,* **51**(1), pp. 69-115.

GLOVER, E., TSIOUTSIOULIKLIS, K., LAWRENCE, S., PENNOCK, D. & FLAKE, G. (2002) Using Web Structure for Classifying and Describing Web Pages. *Proceedings of International World Wide Web Conference WWW2002*, May 7-11, Honolulu, Hawaii, USA

GOODALL, A. H. (2006) Should top universities be led by top researchers and are they?: A citations analysis. *Journal of Documentation*, **62**(3), pp. 388-411.

GOODRUM, A. A., McCAIN, K. W., LAWRENCE, S. & GILES, C. L. (2001) Scholarly Publishing in the Internet Age: A Citation Analysis of Computer Science Literature. *Information Processing and Management*, **37**(5), pp. 661-676.

GRANOVETTER, M. S. (1973) The strength of weak ties. *American Journal of Sociology*, **78**(6), pp. 1360-1380.

GRANOVETTER, M. S. (1983) The strength of weak ties : a network theory revisited. In: Jones, Steve (ed.). *Social structure and network analysis.* Beverly Hills, Cal.: SAGE Publications. pp. 105-130.

GUICE, J. (1998) Looking backward and forward at the Internet. *The Information Society*, **14**, pp. 201-211.

HAAS, S. W. & GRAMS, E. S. (1998) Page and link classifications: connecting diverse resources. *Proceedings of the 3rd ACM Conference on Digital Libraries*. pp. 99-107.

HAAS, S. W. & GRAMS, E. S. (2000) Readers, Authors and Page Structure: A Discussion of Four Questions Arising from a Content Analysis of Web Pages. *Journal of the American Society for Information Science*, **51**(2), pp. 181-192.

HACKETT, S. & PARMANTO, B. (2005) A longitudinal evaluation of accessibility: higher education web sites. *Internet Research,* **15**(3), pp. 281-294.

HAHESSY, J., CARR, F. & FORRESTER, N. (2002) Higher Education Impact Assessment Survey, *Council of Academics and Professional Publishers*. Available: http://www.publishers.org.uk/paweb/paweb.nsf/0/dd103fd25533b94080256caf003d0d95/$FILE/HEImpactAssessmentSurvey.pdf

HARNAD, S. (2006) Online, Continuous, Metrics-Based Research Assessment. Technical Report, ECS, University of Southampton. Deposited 24 Mar 2006.

HARPEL-BURKE, P. (2006) Medium-sized Universities Connect to Their Libraries: Links on University Home Pages and User Group Pages. *Information Technology and Libraries*, **25**(1) pp. 12-23.

HARRIES, G., WILKINSON, D., PRICE, E., FAIRCLOUGH, R. & THELWALL, M. (2004) Hyperlinks as a data source for science mapping. *Journal of Information Science,* **30**(5), pp. 436-447.

HARRISON, C. (2002) Hypertext Links: Whither Thou Goest, and Why. *First Monday,* **7**(10). Available: http://firstmonday.org/issues/issue7_10/harrison/index.html

HARTER, S. & FORD, C. (2000) Web-Based Analysis of E-journal Impact: Approaches, Problems, and Issues. *Journal of the American Society for Information Science*, **51**(13), pp. 1159-1176.

HARTER, S. & KIM, H. (1996) Electronic journals and scholarly communication: a citation and reference study, *Information Research*, **2**(1) paper 9. Available: http://informationr.net/ir/2-1/paper9a.html

HAWKING, D., CRIMMINS, F., CRASWELL, N. & UPSTILL, T. (2004) How valuable is external link evidence when searching enterprise webs? *Fifteenth Australasian Database Conference (ADC'04).*

HEIMERIKS, G. (2006) Discussion of "Operationalising "Websites": lexically, semantically or topologically?" *Cybermetrics*, **10**(1), discussion paper 3. Available: http://www.cindoc.csic.es/cybermetrics/articles/v10i1c3.html

HEIMERIKS, G. & VAN DEN BESSELAAR, P. (2006) Analyzing hyperlinks networks: The meaning of hyperlink based indicators of knowledge production. *Cybermetrics*, **10**(1), paper 1. Available: http://www.cindoc.csic.es/cybermetrics/articles/v10i1p1.html

HEINZE, N. & HU, Q. (2006) The evolution of corporate web presence: A longitudinal study of large American companies. *International Journal of Information Management,* **26**(4), pp. 313-325.

HENZINGER, M. (2000) Link Analysis in Web Information Retrieval. *IEEE Data Engineering Bulletin*, **23**(3), pp. 3-8.

HENZINGER, M. (2001) Hyperlink analysis for the Web. *IEEE Internet Computing*, **5**(1), pp. 45-50.

HENZINGER, M. R., MOTWANI, R. & SILVERSTEIN, C. (2002) Challenges in Web search engines. *SIGIR Forum*, **36**(2), pp. 11−22.

HERNANDEZ-BORGES, A. A., MACIAS-CERVI P. & GASPAR GUADARDO M. A. (1999) Can Examination of WWW Usage Statistics and Other Indirect Quality Indicators Help to Distinguish the Relative Quality of Medical Websites? *Journal of Medical Internet Research*, **1**(1), e1. Available: http://www.jmir.org/1999/1/e1/

HÖRLESBERGER, M. & SCHIEBEL, E. (2006) Web networks of the science system: Weighted hubs and authorities. *Scientometrics*, **66**(2), pp. 263-278.

HUANG, M., CHANG, H. & CHEN, D. (2006) Research evaluation of research-oriented universities in Taiwan from 1993 to 2003. *Scientometrics***,** **67**(3), pp. 419-435.

HUBERMAN, B. A. & ADAMIC, L. A. (1999) Internet: Growth Dynamics of the World Wide Web. *Nature*, **401**, pp. 131.

HUBERMAN, B. A., PIROLLI, P., PITKOW, J. E. & LUKOSE, R. M. (1998) Strong regularities in World Wide Web surfing. *Science,* **280**, pp. 95-97.

HUNTINGTON, P., NICHOLAS, D. & JAMALI, H. R. (2007) Employing log metrics to evaluate search behaviour and success. *Journal of Information Science*, **33**(6).

HUSSEY, J & HUSSEY, R. (1997) Business Research: A Practical Guide for Undergraduate and Postgraduate Students. Macmillan Press Ltd., Basingstoke.

INGWERSEN, P. (1998) The Calculation of Web Impact Factors. *Journal of Documentation*, **54**(2), pp. 236-243.

IVORY, M. Y. & MEGRAW, R. (2005) Evolution of Web Site Design Patterns. ACM *Transactions on Information Systems*, **23**(4), pp. 463–497.

JACOBS, N. (2001) Information technology and interests in scholarly communication: A discourse analysis. *Journal of the American Society for Information Science and Technology*, **52**(13), pp. 1122-1133.

JANSEN, B. J. & SPINK, A. (2005) An analysis of Web searching by European AlltheWeb.com users. *Information Processing & Management*, **41**(2), pp. 361-381.

JANSEN, B. J., SPINK, A. & PEDERSEN, J. (2005) A temporal comparison of AltaVista Web searching. *Journal Of The American Society For Information Science And Technology*, **56**(6), pp. 559–570.

JEPSEN, E. T., SEIDEN, P., INGWERSEN, P. & BJÖRNEBORN, L. (2004) Characteristics of scientific Web publications: Preliminary data gathering and analysis. *Journal of the American Society for Information Science and Technology*, **55**(14), pp. 1239-1249.

KALICZYNSKA, M. (2005) The Polish University Web Sites. *Proceedings of the 10th International Conference on Scientometrics and Informetrics*, *ISSI-2005*. pp. 698-699.

KATZ, J. (2000) Scale-independent Indicators and Research Evaluation. *Science and Public Policy*, **27**(1), pp. 23-26.

KATZ, J. & COTHEY, V. (2006) Web indicators for complex innovation systems. *Research Evaluation*, **15**(2), pp. 85-95.

KHAN, K. & LOCATIS, C. (1998) Searching through Cyberspace: the effects of link display and link density on information retrieval from hypertext on the World Wide Web. *Journal of the American Society for Information Science,* **49**(2), pp. 176-182.

KIM, H. J. (2000) Motivations for hyperlinking in scholarly electronic articles: a qualitative study. *Journal of the American Society for Information Science,* **51**(10), pp. 887-899.

KITCHENS, J.D. & MOSLEY, P.A. (2000) Error 404: or, what is the shelf-life of printed Internet guides? *Library Collections, Acquisitions & Technical Services*, **24**(4), pp. 467-478.

KLEINBERG, J. (1999a) Authoritative Sources in a Hyperlinked Environment. *Journal of the ACM*, **46**(5), pp. 604-632.

KLEINBERG, J. (1999b) The small-world phenomenon: an algorithmic perspective'. Available: http://www.cs.cornell.edu/home/kleinber/swn.d/swn.html

KLEINBERG, J. (2000) Navigation in a small world. *Nature*, **406** pp. 845.

KLEINBERG, J. & LAWRENCE, S. (2001) The structure of the Web. *Science*, **294** pp. 1849-1850.

KLEINBERG, J., KUMAR, R., RAGHAVAN, P., RAJAGOPALAN, S. & TOMKINS, A. (1999) The Web as a graph: measurements, models, and methods'. *Proceedings of the 5th International Computing and Combinatorics Conference.* Also in: *Lecture Notes in Computer Science*, **1627**, pp. 1-18. Available: http://citeseer.nj.nec.com/kleinberg99web.html

KLING, R. & CALLAHAN, E. (2003) Electronic journals, the Internet, and scholarly communication. *Annual Review of Information Science and Technology*, **37**, pp. 127-177.

KLING, R. & MCKIM, G. (2000) Not Just a Matter of Time: Field Differences in the Shaping of Electronic Media in Supporting Scientific Communication. *Journal of the American Society for Information Science*, **51**(14), pp, 1306-1320.

KOEHLER, W. (1999a) Digital libraries and World Wide Web sites and page persistence. *Information research*, **4**(4).

KOEHLER, W. (1999b) An analysis of web page and web site constancy and permanence. *Journal of the American Society for Information Science,* **50**(2), pp. 162-180.

KOEHLER, W. (1999c) Classifying Web sites and Web pages: the use of metrics and URL characteristics as markers. *Journal of Librarianship and Information Science*, **31**(1), pp. 297-307.

KOEHLER, W. (2002) Web Page Change and Persistence - 4 Year Longitudinal web study. *Journal of the American Society for Information Science and Technology*. **53**(2), pp. 162-171.

KOEHLER, W. (2004) A longitudinal study of Web pages continued a consideration of document persistence. *Information research*. **9**(2).

KOUSHA, K. (2005) Webometrics and Scholarly Communication: An Overview. *Quarterly Journal of the National Library of Iran*, (Faslenameh Ketab), **14**(4), pp. 7-16.

KOUSHA, K. & HORRI, A. (2004) The Relationship between Scholarly Publishing and the Counts of Academic Inlinks to Iranian University Web Sites: Exploring Academic Link Creation Motivations. *Journal of Information Management and Scientometrics*, **1**(2), pp. 13-22.

KRETSCHMER, H., KRETSCHMER, U. & KRETSCHMER, T. (2007) Reflection of co-authorship networks in the Web: Web hyperlinks versus Web visibility rates. *Scientometrics*, **70**(2), pp. 519-540

KRETSCHMER, H. & THELWALL, M. (2004) From librametry to webometrics. *Journal of Information Management and Scientometrics*, **1**(1), pp. 1-7.

KUTZ, D. O. & HERRING, S. C. (2005) Micro-Longitudinal Analysis of Web News Updates. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05),* **4**(4) pp. 102.1.

LARSON, R. (1996) Bibliometrics of the World Wide Web: An Exploratory Analysis of the Intellectual Structure of Cyberspace. ASIS 96. Available: http://sherlock.berkeley.edu/asis96/asis96.html

LAWRENCE, S. & GILES, C. L. (1998) Searching the World Wide Web. S*cience,* **280**, pp. 98-100.

LAWRENCE, S. & GILES, C. L. (1999) Accessibility of Information on the Web. *Nature,* **400**, pp. 107-109.

LAWRENCE, S., GILES, C. L. & BOLLACKER, K. (1999) Digital Libraries and Autonomous Citation Indexing. *IEEE Computer*, **32**(6), pp. 67-71.

LEVENE, M. & POULOVASSILIS, A. (2004) Web Dynamics. Adapting to Change in Content, Size, Topology and Use. Berlin Heidelberg: Springer-Verlag.

LEYDESDORFF, L. (2001) *The challenge of scientometrics: the development, measurement, and self-organization of scientific communication*. Parkland, Fla.: Universal Publishers.

LEYDESDORFF, L. & BENSMAN. S. (2005) Classification, power laws, and the logarithmic transformation. Available: http://www.leydesdorff.net/log05/log05.pdf

LEYDESDORFF, L. & CURRAN, M. (2000) Mapping University-Industry-Government Relations on the Internet The Construction of Indicators for a Knowledge-Based Economy. *Cybermetrics*, **4**. Available: http://www.cindoc.csic.es/cybermetrics/articles/v4i1p2.htm

LI, X. (2003) A review of the development and application of the Web Impact Factor, *Online Information Review*, **27**(6), pp. 407-417.

LI, X., THELWALL, M., MUSGROVE, P. & WILKINSON, D. (2003) The relationship between the WIFs or inlinks of Computer Science Departments in UK and their RAE ratings or research productivities in 2001. *Scientometrics*, **57**(2), pp. 239-255.

LI, X., THELWALL, M., MUSGROVE, P. & WILKINSON, D. (2005a) National and international university departmental web site interlinking: Part 1, validation of departmental link analysis. *Scientometrics,* **64**(2)*,* pp. 151-185.

LI, X., THELWALL, M., MUSGROVE, P. & WILKINSON, D. (2005b) National and international university departmental web site interlinking: Part 2, link patterns. *Scientometrics*, **64**(2), pp. 187-208.

LIU, Z. (2004) Perceptions of credibility of scholarly information on the web. *Information Processing and Management*, **40**(6), pp. 1027-1038.

LOKTA, A. J. (1926) The Frequency Distribution of Scientific Distribution. *Journal of the Washington Academy of Science*. **16**(2), pp. 317-323.

MAYFIELD UNIVERSITY CONSULTANTS. (2001). League Tables 2001, *The Times Higher Education Supplement*, May 18, T2-T3.

MCCAIN, K. W. & SALVUCCI, L. J. (2006) How influential is Brooks' law? A longitudinal citation context analysis of Frederick Brooks' The Mythical Man-Month. *Journal of Information Science*, **32**(3), pp. 277-295.

MCMILLAN, S. J. (2001) Survival of the Fittest Online: A Longitudinal Study of Health-Related Web Sites. *Journal of Computer-Mediated Communication,* **6**(3). Available: http://jcmc.indiana.edu/vol6/issue3/mcmillan.html

MEDINA, A., MATTA, I. & BYERS, J. (2000) On the origin of power laws in Internet topologies. *ACM Computer Communications Review*, **30**(2), pp. 18-28.

MENCZER, F. (2001) Links tell us about lexical and semantic Web content. Technical Report Computer Science Abstract CS.IR/0108004. Available: http://arxiv.org/PS_cache/cs/pdf/0108/0108004.pdf

MENDO, F. A. & FITZGERALD, G. (2005) A Preliminary Study of the Evolution of SMEs Web Sites in the UK, *Proceedings of the Ninth European Conference on Software Maintenance and Reengineering*, pp. 375–384.

MIDDLETON, I., McCONNELL, M. & DAVIDSON, G. (1999) Presenting a Model for the Structure and Content of a University World Wide Website. *Journal of Information Science*, **25**(3), pp.219-227.

MILGRAM, S. (1967) The small-world problem. *Psychology Today*, **1**(1), pp. 60-67.

MOLYNEUX, R. E. & WILLIAMS, R. V. (2000) Measuring the Internet. *Annual Review of Information Science and Technology*, **34**, pp. 287-339.

MUSGROVE, P., BINNS, R., KENNEDY, T. & THELWALL, M. (2003) A method for identifying clusters in sets of interlinking Web spaces. *Scientometrics*, **58**(3), pp. 657-672.

NELSON, M. & DOWNIE, J. (2002) Informetric analysis of a music database. *Scientometrics*, **54**(2), pp. 243-255.

NICHOLSON, S. (2006) The basis for bibliomining: Frameworks for bringing together usage-based data mining and bibliometrics through data warehousing in digital library services. *Information Processing and Management*, **42**(3), pp. 785-804.

NOBLE PUBLISHING CO. (1999) *Noble's Higher Education Financial Yearbook 1999*. Noble Publishing: Edinburgh.

NORUZI, A. (2006a) Web presence and impact factors for Middle-Eastern countries. *Online*, **30**(2), pp. 22-28.

NORUZI, A. (2006b) The Web Impact Factor: a critical review. *The Electronic Library,* **24**(4), pp. 490-500.

NTOULAS, A., CHO, J. & OLSTON, C. (2004) What's new on the web? The evolution of the web from a search engine perspective. *In Proceedings of the 13th International World Wide Web Conference, May 17-22, 2004.*

ONYANCHA, O, B. & OCHOLLA, D, N. (2007) The Performance of South African and Kenyan Universities on the World Wide Web: a Web Link Analysis. *Cybermetrics*, **11**(1), Paper 2. Available: http://www.cindoc.csic.es/cybermetrics/articles/v11i1p2.html

OPPENHEIM, C. & STUART, D. (2004) Is there a correlation between investment in an academic library and a higher educations ratings in the Research Assessment Exercise? *Aslib Proceedings*, **56**(3), pp. 156-165.

ORTEGA, J.L., AGUILLO, I.F. & PRIETO, J.A. (2006) Longitudinal Study of Contents and Elements in the Scientific Web environment. *Journal of Information Science*, **32**(4), pp. 344-351.

PARK, H, W. (2002) Examining the determinants of who is hyperlinked to whom: a survey of webmasters in Korea. *First Monday*, **7**(11). Available: http://firstmonday.org/issues/issue7_11/park/

PARK, H. & THELWALL, M. (2003) Hyperlink Analysis of the World Wide Web: A Review. *Journal of Computer-Mediated Communication* **8**(4). Available: http://www.ascusc.org/jcmc/vol8/issue4/park.html

PARK, H. & THELWALL, M. (2006) Web science communication in the age of globalization: Links among universities' websites in Asia and Europe. *New Media & Society*, **8**(4), pp. 629-650.

PARK, H. W., BARNETT, G. A. & NAM, I. (2002) Hyperlink-affiliation network structure of top web sites: Examining affiliates with hyperlink in Korea. *Journal of the American Society for Information Science and Technology*, **53**(7), pp. 592-601.

PAYNE, N. & THELWALL, M. (2004) A Statistical Analysis of UK Academic Web Links. *Cybermetrics*, **8**(1), paper 2. Available: http://www.cindoc.csic.es/cybermetrics/articles /v8i1p2.html

PAYNE, N. & THELWALL, M. (2005) Mathematical Models for Academic Webs: Linear Relationship or Non-Linear Power Law? *Information Processing and Management*, **41**(6), pp. 1495-1510.

PAYNE, N. & THELWALL, M. (2007a) A Longitudinal Study of Academic Webs: Growth and Stabilisation. *Scientometrics*, **71**(3), pp. 523-539.

PAYNE, N. & THELWALL, M. (2007b, to appear) Longitudinal Trends in Academic Web Links. *Journal of Information Science*.

PAYNE, N. & THELWALL, M. (2007c, to appear) A Longitudinal Analysis of Alternative Document Models. *ASLIB Proceedings.*

PENNOCK, D., FLAKE, G., LAWRENCE, S., GLOVER, E. & GILES, C. L. (2002) Winners Don't Take All: Characterizing the Competition for Links on the Web. *Proceedings of the National Academy of Sciences*, **99**(8), pp. 5207-5211.

PHILIPS, E. M. (1992) The PhD: assessing quality at different stages of its development, in O. Zuber-Skerritt (ed.) *Starting Research: Supervision and Training*. Brisbane, Queensland: Tertiary Education Institute, University of Queensland.

PINTO, M., BERROCAL, J. L. A., GARCÍA, J. A. C., MARCIAL, V. F., FIGUEROLA, C. G., MARCO, J. G., GÓMEZ, C. C. & ZAZO, R. A. (2005) Quality assessment of Spanish universities' web sites focused on the European Research Area. *Scientometrics*, **65**(1), pp. 67-93.

PIROLLI, P. L.T. & PITKOW, J. E. (1999) Distributions of surfers' paths through the World Wide Web. *World Wide Web*, **2**, pp. 29-45.

POLANCO, X, BOUDOURIDES, M. A., BESAGNI, D. & ROCHE, I. (2001) *Clustering and mapping Web sites for displaying implicit associations and visualising networks*. Patras, Greece: University of Patras. Available: http://www.math.upatras.gr/~mboudour/articles/Web_clustering&mapping.pdf

PRIME, C., BASSECOULARD, E. & ZITT, M. (2002) Co-citations and co-sitations: a cautionary view on an analogy. *Scientometrics*, **54**(2), pp. 291-308.

REDNER, S. (1998) How popular is your paper? An empirical study of the citation distribution. *European Physical Journal B*, 4: 131-134. Preprint available: http://arxiv.org/PS_cache/cond-mat/pdf/9804/9804163.pdf

ROBERTS, P. (1999) Scholarly Publishing, Peer Review and the Internet. *First Monday*, **8**(2). Available: http://firstmonday.org/issues/issue8_2/roberts/index.html

RODRIGUEZ I GAIRÍN, J. M. (1997). 'Valorando el impacto de la informacion en Internet: Alta-vista, el "Citation Index" de la Red'. *Revista Espanola deocumentacion Scientifica*, **20**(2), pp. 175-181.

ROSENBAUM, H. (1998). Web-based Community Networks: A Study of Information Organization and Access. *Proceedings of the 61st Annual Meeting of the American Society for Information Science,* **35**, pp. 516-530. Medford, NJ: Information Today, Inc.

ROUSSEAU, R. (1997) Situations: an Exploratory Study. *Cybermetrics*, **1**(1), paper 1. Available: http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html

ROUSSEAU, R. (1998/1999). Daily time series of common single word searches in AltaVista and NorthernLight. *Cybermetrics,* **2/3**(1) Available: http://www.cindoc.csic.es/cybermetrics/articles/v2i1p2.html

ROUSSEAU, R. (1999) Time Evolution of the Number of Hits in Keyword Searches on the Internet, *Post Conference Seminar – Cybermetrics '99 at the Seventh International Conference on Scientometrics and Informetrics*, July 9 1999, Colima, Mexico. Available: http://www.cindoc.csic.es/cybermetrics/cybermetrics99.html

ROUSSEAU, R. (2001) Evolution in time of the number of hits in keyword searches on the Internet during one year, with special attention to the use of the word *euro. Proceedings of the 8th International Conference on Scientometrics & Informetrics, ISSI-2001*. pp. 619-627.

ROUSSEAU, B. & ROUSSEAU, R. (2000) LOTKA: A Program to fit a Power Law Distribution to Observed Frequency Data. *Cybermetrics*, **4**(1), paper 4. Available: http://www.cindoc.csic.es/cybermetrics/articles/v4i1p4.html

ROUSSEAU, R. & THELWALL, M. (2004) Escher staircases on the world wide web. *FirstMonday,* **9**(6). Available: http://www.firstmonday.org/issues/issue9_6/rousseau/index.html

SCHARNHORST, A. & WOUTERS, P. (2006) Web indicators – a new generation of S&T indicators? *Cybermetrics*, **10**(1), paper 6. Available: http://www.cindoc.csic.es/cybermetrics/articles/v10i1p6.html

SEBASTIANI, F. (2002) Machine learning in automated text categorization. *ACM Computing Surveys*, **34**(1), pp. 1–47.

SHEPHERD, M. A., WATTERS, C. R. & CAI, Y. (1990) Transient hypergraphs for citation networks. *Information Processing & Management*, **26**(3), pp. 395-412.

SHI, Y. (2006) E-Government Web Site Accessibility in Australia and China: A Longitudinal Study. *Social Science Computer Review*, **24**(3), pp. 378-385.

SMALL, H. (1973) Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science,* **24**(4), pp. 265-269.

SMITH, A. (1999) A Tale of Two Web Spaces: Comparing Sites Using Web Impact Factors. *Journal of Documentation*, **55**(5), pp. 577-592.

SMITH, A. (2004) Web links as analogues of citations. *Information Research*. **9**(4).

SMITH, A. & THELWALL, M. (2001) Web impact factors and university research links. *Proceedings of the 8th International Conference on Scientometrics and Informetrics*, *ISSI-2001*. pp. 657-664.

SMITH, A. & THELWALL, M. (2002) Web Impact Factors for Australasian Universities. *Scientometrics*, **54**(3), pp. 363-380.

SNYDER, H. & ROSENBAUM, H. (1999) Can search engines be used as tools for web-link analysis? A critical view. *Journal of Documentation*, **55**(4), pp. 375-384.

SPINK, A., WOLFRAM, D., JANSEN, B. J. & SARACEVIC, T. (2001) Searching the Web: The public and their queries. *Journal of the American Society for Information Science and Technology*, **53**(2), pp. 226-234.

SRIDHAR, M. S. (2007) Research Methodology Part 3: Research Design and Plan. *E-LIS Library Instructional Material*. Available: http://eprints.rclis.org/archive/00010170/

STUART, D. & THELWALL, M. (2005) What can university-to-government web links tell us about a university's research productivity and the collaborations between universities and government? *Proceedings of the 10th International Conference on Scientometrics and Informetrics*, *ISSI-2005*. pp. 188-192.

STUART, D., THELWALL, M. & HARRIES, G. (2006) UK academic web links and collaboration – an exploratory study. *Journal of Information Science*, **33**(2), pp. 231-246.

SUNSTEIN, C. R. (2001) *Republic.com*. Princeton, NJ: Princeton University Press.

TAGUE-SUTCLIFFE, J. (1992) An introduction to informetrics. *Information Processing & Management*, **28**(1), pp. 1-3.

TANG, R. & THELWALL, M. (2003) U.S. academic departmental Web-site interlinking in the United States Disciplinary differences. *Library & Information Science Research*, **25**, pp. 437-458.

TANG, R. & THELWALL, M. (2004) Patterns of national and international web inlinks to US academic departments: An analysis of disciplinary variations. *Scientometrics,* **60**(3), pp. 475-485.

TANG, R. & THELWALL, M. (2005, to appear). A hyperlink analysis of US public and academic libraries' Web sites. *Library Quarterly*.

THELWALL, M. (2000) Web Impact Factors and Search Engine Coverage. *Journal of Documentation*, **56**(2), pp. 185-189.

THELWALL, M. (2001a) A Web crawler design for data mining. *Journal of Information Science*, **27**(5), pp. 319-325.

THELWALL, M. (2001b) A Publicly Accessible Database of UK University Website Links and a Discussion of the Need for Human Intervention in Web Crawling. *University of Wolverhampton.* Available: http://www.scit.wlv.ac.uk/~cm1993/papers/a_publicly_accessible_ database.pdf

THELWALL, M. (2001c) Exploring the link structure of the Web with network diagrams. *Journal of Information Science,* **27**(6), pp. 393-401.

THELWALL, M. (2001d) Extracting Macroscopic Information from Web Links. *Journal of the American Society for Information Science and Technology*, **52**(13), pp. 1157-1168.

THELWALL, M. (2001e) The Responsiveness of Search Engine Indexes. C*ybermetrics*, **5**. Available: http://www.cindoc.csic.es/cybermetrics/articles /v5i1p1.htm

THELWALL, M. (2001f) Results from a web impact factor crawler. *Journal of Documentation*, **57**(2), pp. 177-191.

THELWALL, M. (2002a) A comparison of sources of links for academic Web impact factor calculations. *Journal of Documentation*, **58**(1), pp. 66-78.

THELWALL, M. (2002b) An Initial Exploration of the Link Relationship Between UK University Websites. *ASLIB Proceedings*, **54**(2), pp. 118-126.

THELWALL, M. (2002c) Can Google's PageRank be used to find the most important academic web pages. *Journal of Documentation*, **59**(2), pp. 205-217.

THELWALL, M. (2002d) A Research and Institutional Size Based Model for National University Website Interlinking. *Journal of Documentation*, **58**(6), pp. 683-694.

THELWALL, M. (2002e) Conceptualising Documentation on the Web: An Evaluation of Different Heuristic-Based Models for Counting Links Between University Websites. *Journal of the American Society for Information Science and Technology,* **53**(12), pp. 995-1005.

THELWALL, M. (2002f) Evidence for the Existence of Geographic Trends in University Website Interlinking. *Journal of Documentation,* **58**(5), pp. 563-574.

THELWALL, M. (2002g) Research dissemination and invocation on the web. *Online Information Review*, **26**(6), pp. 413-420.

THELWALL, M. (2002h) Research Note: in praise of Google: finding law journal web sites. *Online Information Review*, **26**(4), pp. 271-272.

THELWALL, M. (2002i) The top 100 linked pages on UK university Web sites: high inlink counts are not usually directly associated with quality scholarly content. *Journal of Information Science*, **28**(6), pp. 485-491.

THELWALL, M. (2002/3) A Free Database of University Web Links Data Collection Issues. *Cybermetrics*, 6/7. Available: http://www.cindoc.csic.es/cybermetrics/articles/v6i1p2.htm

THELWALL, M. (2003a) A layered approach for investigating the topological structure of communities in the web. *Journal of Documentation*, **59**(4), pp. 410-429.

THELWALL, M. (2003b) Web use and peer interconnectivity metrics for academic Web sites. *Journal of Information Science*, **29**(1), pp. 1-10.

THELWALL, M. (2003c) What is this link doing here Beginning a fine-grained process of identifying reasons for academic hyperlink creation. *Information Research*. **8**(3).

THELWALL, M. (2004a) Can the Web give useful Information about commercial uses of scientific research? *Online Information Review*, **28**(2), pp. 120-130.

THELWALL, M. (2004b) Methods for reporting on the targets of links from national systems of university Web sites. *Information Processing and Management*, **40**(1), pp. 125-144.

THELWALL, M. (2004c) *Link Analysis: An Information Science Approach*. Elsevier Academic Press.

THELWALL, M. (2005) Scientific Web Intelligence: Finding relationships in university webs. *Communications of the ACM,* **48**(7), pp. 93-96.

THELWALL, M. (2006a) Interpreting social science link analysis research: A theoretical framework. *Journal of the American Society for Information Science and Technology*, **57**(1), pp. 60-68.

THELWALL, M. (2006b). Bloggers during the London attacks: Top information sources and topics. *WWW2006 blog workshop*. Available: http://www.blogpulse.com/www2006-workshop/papers/blogs-during-london-attacks.pdf

THELWALL, M. (2007/8, to appear) Extracting accurate and complete results from search engines: Case study Windows Live. *Journal of the American Society for Information Science and Technology*.

THELWALL, M. & AGUILLO, I. (2003) La salud de las Web universitarias españolas. *Revista Española de Documentación Científica,* **26**(3), pp. 291-305.

THELWALL, M., BARJAK, F. & KRETCHMER, H. (2006) Web links and gender in science: An exploratory analysis. *Scientometrics*, **67**(3), pp. 373-383.

THELWALL, M., BINNS, R., HARRIES, G., PAGE-KENNEDY, T., PRICE L. & WILKINSON, D. (2001) Custom Interfaces for advanced queries in search engines, *ASLIB Proceedings*, **53**(10), pp. 413-422.

THELWALL, M., BINNS, R., HARRIES, G., PAGE-KENNEDY, T., PRICE, L. & WILKINSON, D. (2002) European Union associated University Websites. *Scientometrics*, **53**(1), pp. 95-111.

THELWALL, M. & HARRIES, G. (2003) The Connection between the Research of a University and Counts of Links to its Web Pages: An Investigation Based Upon a Classification of the Relationships of Pages to the Research of the Host University. *Journal of the American Society for Information Science and Technology*, **54**(7), pp. 594-602.

THELWALL, M. & HARRIES, G. (2004a) Do better scholars' web publications have significantly higher online impact? *Journal of the American Society for Information Science and Technology*, **55**(2), pp. 149-159.

THELWALL, M. & HARRIES, G. (2004b) Can personal web pages that link to universities yield information about the wider dissemination of research? *Journal of Information Science*, **30**(3), pp. 243-256.

THELWALL, M., HARRIES, G. & WILKINSON, D. (2003) Why do Web sites from different academic subjects interlink? *Journal of Information Science*, **29**(6), pp. 453-471.

THELWALL, M., LI, X., BARJAK, F. & ROBINSON, S. (2007, to appear). Assessing the web connectivity of research groups on an international scale. *ASLIB Proceedings*.

THELWALL, M. & PAYNE, N. (2005) Link Analysis: An Informetric Technique. *Proceedings of the 10th International Conference on Scientometrics and Informetrics*, *ISSI-2005*. pp. 681-682.

THELWALL, M. & PRICE, E. (2003) Disciplinary differences in academic web presence – A statistical study of the UK. *Libri*, **53**(4), pp. 242-253.

THELWALL, M. & SMITH, A. (2002) Interlinking between Asia-Pacific University Web Sites. *Scientometrics*, **55**(3), pp. 363-376.

THELWALL, M. & TANG, R. (2003) Disciplinary and linguistic considerations for academic Web linking: An exploratory hyperlink mediated study with Mainland China and Taiwan. *Scientometrics*, **58**(1), pp. 155-181

THELWALL, M., TANG, R. & PRICE, L. (2003) Linguistic patterns of academic Web use in Western Europe. *Scientometrics*, **56**(3), pp. 417-432.

THELWALL, M. & VAUGHAN, L. (2004a) New versions of PageRank employing alternative web documents models. *ASLIB Proceedings*, **56**(1), pp. 24-33.

THELWALL, M. & VAUGHAN, L. (2004b) A fair history of the Web? Examining country balance in the Internet Archive. *Library & Information Science Research*, **26**(2), pp. 162-176.

THELWALL, M. VAUGHAN, L. & BJÖRNEBORN, L. (2005) Webometrics. *Annual Review of Information Science and Technology*, **39**, pp. 81-135.

THELWALL, M., VAUGHAN, L., COTHEY, V., LI, X. & SMITH, A. (2003) Which academic subjects have most online impact? A pilot study and a new classification process, *Online Information Review*, **27**(5), pp. 333-343.

THELWALL, M. & WILKINSON, D. (2003a) Finding Similar Academic Websites with Links, Bibliometric Couplings and Colinks. *Information Processing and Management*. **40**(3), pp. 515-526.

THELWALL, M. & WILKINSON, D. (2003b) Three Target Document Range Metrics for University Websites. *Journal of the American Society for Information Science and Technology*, **54**(6), pp. 489-496.

THELWALL, M. & WILKINSON, D. (2003c) Graph Structure in Three National Academic Webs: Power Laws with Anomalies. *Journal of the American Society for Information Science and Technology*, **54**(8), pp. 706-712.

THOMAS, O. & WILLET, P. (2000) Webometric Analysis of Departments of Librarianship and Information Science. *Journal of Information Science*, **26**(6), pp. 421-428.

THOMS, L. & THELWALL, M. (2005) Academic home pages: Reconstruction of the self. *First Monday*, **10**(12). Available: http://www.firstmonday.org/issues/issue10_12/thoms/

TURNBULL, D. (1996) Bibliometrics and the World Wide Web, (Technical Report FIS-12-19-1996-1). Faculty of Information Studies, University of Toronto. Available: http://www.ischool.utexas.edu/~donturn/research/bibweb.html

UBERTI T.E. & MAGGIONI M.A., (2004) *Infrastrutture ICT e relazionalità potenziale. Un esercizio di "hyperlinks counting" a livello sub-nazionale*, Quaderno DISEIS n. 0402, Università Cattolica del sacro Cuore, Milano, Vita e Pensiero.

UPSTILL, T., CRASWELL, N. & HAWKING, D. (2003) Query-independent evidence in home page finding. *ACM Transactions on Information Systems,* **21**(3), pp. 286-313.

UZUN, A. (2006) Statistical relationship of some basic bibliometric indicators in Scientometrics research. *Proceedings of the International Workshop on Webometrics, Informetrics and Scientometrics*, pp. 87-91, 10-12 May 2006, Nancy-France.

VAN RAAN, A. (2001) Bibliometrics and internet: some observations and expectations. *Scientometrics*, **50**(1), pp. 59-63.

VASILEIADOU, E. & VAN DEN BESSELAAR, P. (2006) Linking shallow, linking deep. How scientific intermediaries use the Web for their network of collaborators. Cybermetrics, **10**(1), paper 4. Available: http://www.cindoc.csic.es/cybermetrics/articles/v10i1p4.html

VAUGHAN, L. (2001) Statistical Methods for the Information Professional: A Practical, Painless Approach to Understanding, Using, and Interpreting Statistics, *ASIST Monograph Series.* Medford, New Jersey: Information Today, Inc.

VAUGHAN, L. & HYSEN, K. (2002) Relationship between links to journal Web sites and impact factors. *Aslib Proceedings,* **54**(6), pp. 356-361

VAUGHAN, L. & SHAW, D. (2003) Bibliographic and web citations: What is the difference? *Journal of the American Society for Information Science and Technology*, **54**(14), pp. 1313-1322.

VAUGHAN, L. & THELWALL, M. (2003) Scholarly use of the Web: what are the key inducers of links to journal Web sites? *Journal of the American Society for Information Science and Technology,* **54**(1), pp. 29-38.

VAUGHAN, L. & THELWALL, M. (2004) Search engine coverage bias: Evidence and possible causes. *Information Processing and Management,* **40**(4), pp. 693-707.

VAUGHAN, L. & THELWALL, M. (2005) A modelling approach to uncover hyperlink patterns the case of Canadian universities. *Information Processing and Management*, **41**(2), pp. 347-359.

VAUGHAN, M. W. & DILLON, A. (2000) Learning the shape of information: a longitudinal study of Web-news reading. *Proceedings of the fifth ACM conference on Digital libraries*, pp. 236-237.

VAUGHAN, M. W. & DILLON, A. (2006) Why structure and genre matter for users of digital information: A longitudinal experiment with readers of a web-based newspaper. *International Journal of Human Computer Studies*, **64**(6), pp. 502-526.

WALKER, J. (2002) Links and power: the political economy of linking on the Web. *Proceedings of Hypertext 2002*. Baltimore: ACM Press. pp. 78-79.

WANG, P., BERRY, M. W. & YANG, Y. (2003) Mining longitudinal web queries: Trends and patterns. *Journal of the American Society for Information Science and Technology,* **54**(8), pp. 743-758.

WATTS, D. J. (1999a) The Internet, the small world, and the nature of distance. Messages, Museum of Science, Boston. Available: http://aries.mos.org/internet/essay.html

WATTS, D. J. (1999b) Networks, dynamics, and the small-world phenomenon. *American Journal of Sociology*, **105**(2), pp. 493-527.

WATTS, D. J. & STROGATZ, S. H. (1998) Collective Dynamics of 'Small-World' Networks. *Nature,* **393**, pp. 440-442

WEARE, C. & LIN, W. (2000) Content analysis of the World Wide Web: opportunities and challenges. *Social Science Computer Review*, **18**(3), pp. 272-292.

WILKINSON, D., HARRIES, G., THELWALL, M. & PRICE, E. (2003) Motivations for academic Web site interlinking: Evidence for the Web as a novel source of information on informal scholarly communication. *Journal of Information Science*, **29**(1), pp. 49-56.

WILKINSON, D., THELWALL, M. & LI, X. (2003) Exploiting hyperlinks to study academic Web use. *Social Science Computer Review*, **21**(3), pp. 340-351.

WOLFRAM, D. (1992) Applying informetric characteristics of databases to IR system file design, Part I: informetric models. *Information Processing and Management*, **28**(1), pp. 121-133.

WOLFRAM, D. (2000) Applications of Informetrics to IR research. *Informing Science*, **3**(2).

WORMELL, I. (1998) Informetrics: an emerging subdiscipline in information science. *Asian Libraries*, **7**(10), pp. 257-268.

WOUTERS, P., HELLSTEN, I. & LEYDESDORFF, L. (2004) Internet time and the reliability of search engines. *First Monday*, **9**(10). Available: http://firstmonday.org/issues/issue9_10/wouters/index.html

WYNAR, B. S. (1971) *Research Methods in Library Science*. University of Michigan: Libraries Unlimited.

YANG, Y., SLATTERY, S. & GHANI, R. (2002) A Study of Approaches to Hypertext Categorization. *Journal of Intelligent Information Systems*, **18** (2-3), pp. 219-241.

YEUNG, W. L. & LU, M. (2004) Functional characteristics of commercial web sites: a longitudinal study in Hong Kong. *Information & Management*, **41**(4), pp. 483-495.

YOOK, S., JEONG, H. & BARABÁSI, A. (2002) Modeling the Internet's large-scale topology. *Proceedings of the National Academy of Sciences*, **99**(21), pp. 13382-13386.

YUAN, W. (1997) End-User Searching Behavior in Information Retrieval: A Longitudinal Study. *Journal of the American Society for Information Science,* **48**(3), pp. 218-234.

ZHANG, Y. (2001) Scholarly use of Internet-based electronic resources. *Library Trends*, **47**(4), pp. 746-770.

Glossary

ADM – Alternative Document Model (originally known as Advanced Document Model).

Bibliometrics – the study of the quantitative aspects of the production, dissemination and use of recorded information. It develops mathematical models and measures for these processes and then uses these models and measures for prediction and decision making.

Citation – a reference by one publication of another. A citation is the reference viewed from the perspective of the referenced document.

Co-linked – two pages that both have inlinks from a third page are co-linked.

Co-linking –two pages that both have outlinks to a third page are co-linking (sometimes also described as bibliometric coupling or just coupling).

Cybermetrics – the study of the quantitative aspects of the construction and use of information resources, structures and technologies on the whole Internet, drawing on bibliometric and informetric approaches.

HTML – HyperText Markup Language. The coding language in which web pages are described. This is interpreted by web browsers to produce the web pages that web users see, and is processed by web crawlers to extract the embedded links.

HTTP – HyperText Transfer Protocol. The mechanism used by programs such as web browsers and crawlers to communicate with a web server, for example to request a web page.

Hyperlink – a feature in a web page that allows users to click to navigate to a different web page (or a different part of the same page). Hyperlinks are also called links and can also be found in hypertext environments other than the web.

Informetrics – the investigation of quantitative aspects of information processes, particularly those using text; it is the quantitative arm of Information Science and of Library Science. Informetrics incorporates the older field of Bibliometrics and the new areas of Cybermetrics and Webometrics.

Inlinks – incoming links to web document sets (e.g. pages, domains, sites) from other web document sets. Can be considered to be analogous with the term ‘citation’ in bibliometrics.

Interlink –normally a link between two different web sites, also referred to as an inter-site link.

Internet – a worldwide, publicly accessible network of interconnected computer networks that transmit data by packet switching using the standard Internet Protocol (IP).

ISI – Institute of Scientific Information

JIF – Journal Impact Factor

LIS – Library and Information Science

Outlinks – outgoing links from web document sets (e.g. pages, domains, sites) to other web document sets. Can be considered to be analogous with the term ‘reference’ in bibliometrics.

RAE – Research Assessment Exercise

Scientometrics – the investigation of quantitative aspects of science; it is the quantitative arm of the Science of Science, of Scientific Communication Studies and of Science Policy Studies. Scientometrics and Informetrics are bound through their mutual interest in scientific literature. Scientometrics involves quantitative studies of scientific activities, including publications, and so overlaps bibliometrics to some extent.
Selflink –a link from a web page to the same page, perhaps to a different part of the page. If qualified by a web unit, this implies that the link should target a page inside of the specified unit including domains, sites and universities. For example a site selflink is a link from any page in a site to any page in the same site. Site selflink is synonymous with 'internal site link', or sometimes just 'internal link'.
Site Links – points from one area of the document set in question to another area within the same set. Are also known as 'internal links' and can be at various levels.
TLD – Top Level Domain
Web (World Wide Web) – a system of interlinked, hypertext documents accessed via the Internet.
Webometrics – the quantitative study of web-related phenomena. The study of the quantitative aspects of the construction and use of information resources, structures and technologies on the web, drawing on bibliometric and informetric approaches.
Web Site – a related collection of web pages containing a home page, which is the first document users see when they enter the site. The site might also contain additional documents and files. Each site is owned and managed by an individual, company or organization.
Web Space – a collection of related web pages, (sub)domains or web sites.
WIF – Web Impact Factor

Index

Appendix 1: Source and Target Domain, Directory and Page ADM Data for 111 UK Universities June/July 2002

University Name	Domain Name	Source Domains	Target Domains	Source Directories	Target Directories	Source Pages	Target Pages
University of Aberdeen	abdn.ac.uk	67	19975	8576	42223	102607	193678
University of Wales, Aberystwyth	aber.ac.uk	41	16504	6605	35068	74963	120183
Anglia Polytechnic University	anglia.ac.uk	56	5374	1239	9133	14711	27570
Aston University	aston.ac.uk	60	8052	1195	11808	12857	27046
University of Wales Bangor	bangor.ac.uk	62	5093	2220	9757	17370	32014
University of Bath	bath.ac.uk	32	14470	6276	31320	51904	93839
Bath Spa University College	bathspa.ac.uk	10	1734	505	2838	2014	5069
University of Birmingham	bham.ac.uk	302	23612	16012	60589	260225	369203
University of Bournemouth	bournemouth.ac.uk	42	6873	1700	11101	23030	37026
University of Bradford	brad.ac.uk	45	11072	3548	22180	29697	63507
University of Bristol	bris.ac.uk	192	24552	10569	54944	113732	234005
Oxford Brookes University	brookes.ac.uk	27	9670	2382	16008	23278	41212
Brunel University	brunel.ac.uk	22	9998	7650	22952	50706	78466
University of Brighton	bton.ac.uk	64	14015	2687	25287	49086	81880
University of Buckingham	buckingham.ac.uk	3	433	210	745	1430	2155
University of Cambridge	cam.ac.uk	614	40929	24296	112466	205847	402888
Canterbury Christ Church University College	cant.ac.uk	48	3700	1634	7243	17067	25198
Cardiff University	cf.ac.uk	63	16886	5430	31589	44345	91950
Chichester College	chichester.ac.uk	4	29	36	65	211	326
City University, London	city.ac.uk	39	8366	5207	18471	66336	169527
Cranfield University	cranfield.ac.uk	42	5660	1295	11050	9509	26978
University of Coventry	coventry.ac.uk	22	4593	816	7130	9691	21417
University of Derby	derby.ac.uk	30	2882	1260	5407	11841	22564
De Montfort University	dmu.ac.uk	77	13720	7279	29904	59917	101670
University of Dundee	dundee.ac.uk	58	8858	3737	17846	51316	77154
University of Durham	dur.ac.uk	51	14225	7731	32692	76893	131184
University of Edinburgh	ed.ac.uk	350	38250	19390	98705	229093	462984

University of Essex	essex.ac.uk	60	18974	4971	37605	52329	108656
University of Exeter	ex.ac.uk	51	24515	6617	46421	73500	143515
Glasgow Caledonian University	gcal.ac.uk	60	3730	645	6044	5551	17428
University of Glasgow	gla.ac.uk	372	31091	20790	84293	366924	499546
University of Glamorgan	glam.ac.uk	22	4026	1365	7132	14626	22175
University of Gloucestershire	glos.ac.uk	2	624	65	902	1488	2557
Goldsmiths College, University of London	goldsmiths.ac.uk	26	5432	1064	9146	9989	21879
University of Greenwich	gre.ac.uk	39	7784	2424	15071	21483	43982
Harper-Adams Agricultural College	harper-adams.ac.uk	1	158	79	244	520	919
University of Hertfordshire	herts.ac.uk	45	7775	3086	16572	63319	113728
University of Huddersfield	hud.ac.uk	23	3720	1331	6753	10918	18381
University of Hull	hull.ac.uk	41	6100	1406	10875	15619	78399
Heriot-Watt University	hw.ac.uk	61	18079	5993	35037	60372	109750
Imperial College, University of London	ic.ac.uk	241	19422	17850	67679	177208	301373
King’s College London	kcl.ac.uk	61	13181	5811	29149	59037	101713
University of Keele	keele.ac.uk	14	19233	2066	30630	16809	54419
Kingston University	king.ac.uk	37	4045	1215	6997	13557	22683
University of Wales, Lampeter	lamp.ac.uk	7	2138	316	3501	3497	7886
University of Lancaster	lancs.ac.uk	89	19042	7223	40785	71512	173110
University of Loughborough	lboro.ac.uk	84	9948	5672	25859	34202	64387
University of Leicester	le.ac.uk	35	11810	3605	24973	34491	73503
University of Leeds	leeds.ac.uk	186	25542	14247	64866	147827	293725
London Guildhall University	lgu.ac.uk	6	2720	1227	6711	13478	21056
University of Liverpool	liv.ac.uk	28	5862	1954	11448	18262	34652
Liverpool John Moores University	livjm.ac.uk	33	7433	2994	13865	30525	46690
Leeds Metropolitan University	lmu.ac.uk	19	1701	741	3017	10490	14434
London School of Economics	lse.ac.uk	53	4747	3362	11719	29757	46529
University of Luton	luton.ac.uk	8	186	610	1020	1338	4934
University of Manchester	man.ac.uk	248	19508	7686	44282	124070	219145
University of Middlesex	mdx.ac.uk	55	9126	3773	22527	41933	90331
Manchester Metropolitan University	mmu.ac.uk	67	9207	2006	16669	21108	43656

Napier University	napier.ac.uk	54	14022	10388	34626	98712	168827
University of Newcastle	ncl.ac.uk	144	23209	12455	54271	116901	194415
University of Wales College Newport	newport.ac.uk	55	1843	509	2728	2949	5667
Northampton University College	northampton.ac.uk	8	4861	753	7062	6666	15021
University of Northumbria	northumbria.ac.uk	50	4042	1086	7140	13123	26546
University of Nottingham	nott.ac.uk	115	15769	6526	33342	70036	116427
Nottingham Trent University	ntu.ac.uk	73	9066	2222	16019	19491	40273
Open University	open.ac.uk	194	13002	5984	32493	45584	105042
University of Oxford	ox.ac.uk	521	35533	18030	89435	197162	355301
University of Paisley	paisley.ac.uk	2	66	11	81	122	198
University of Plymouth	plym.ac.uk	85	8338	3083	16030	23703	47164
University of Portsmouth	port.ac.uk	78	9553	4183	19055	69988	100335
Queen Margaret University College	qmced.ac.uk	24	6006	1519	10357	8794	26451
Queen Mary, University of London	qmw.ac.uk	66	5928	2233	11626	23303	56097
Queen’s University Belfast	qub.ac.uk	80	14781	6079	31133	54601	100652
University of Reading	rdg.ac.uk	109	12472	6732	27776	91278	136561
The Robert Gordon University	rgu.ac.uk	18	5388	1164	9107	7589	18630
Royal Holloway, University of London	rhul.ac.uk	68	8714	6239	20552	42735	66987
University of Salford	salford.ac.uk	90	2547	1703	5331	8246	15624
South Bank University	sbu.ac.uk	51	12492	1632	21517	19067	47295
University of Sheffield	shef.ac.uk	67	10557	6424	24356	44099	71132
Sheffield Hallam University	shu.ac.uk	40	7886	4867	17930	56375	84127
School of Oriental and African Studies	soas.ac.uk	15	1187	678	2385	5681	8182
University of Southampton	soton.ac.uk	305	27726	14150	67016	121605	244775
University of St Andrews	st-and.ac.uk	38	7025	2587	13111	35620	54901
University of Staffordshire	staffs.ac.uk	126	9023	9801	26099	83747	136334
University of Stirling	stir.ac.uk	93	6971	3511	14465	24153	40212
University of Strathclyde	strath.ac.uk	171	15321	9663	38018	80476	167587
University of Sunderland	sunderland.ac.uk	57	8667	2105	16080	31276	53739
The Surrey Institute of Art and Design University College	surrart.ac.uk	1	78	36	119	343	460

University of Surrey	surrey.ac.uk	46	13345	4384	26630	89091	123135
University of Sussex	susx.ac.uk	73	14135	6030	30290	68017	117356
University of Wales Swansea	swan.ac.uk	65	14407	2751	24688	27521	57175
University of Abertay, Dundee	tay.ac.uk	18	306	168	555	1131	1761
University of Teeside	tees.ac.uk	31	2366	1042	4103	11174	17820
Thames Valley University	tvu.ac.uk	21	2223	854	3737	7214	11171
University of Central England	uce.ac.uk	32	3294	2474	8207	21317	33465
University College London	ucl.ac.uk	229	24378	13283	60195	142017	222723
University of Central Lancashire	uclan.ac.uk	10	6208	1208	9577	12561	23480
University of East Anglia	uea.ac.uk	54	16611	3991	29968	46966	83743
University of East London	uel.ac.uk	14	5518	1916	10320	17657	31152
University of Kent at Canterbury	ukc.ac.uk	32	9134	5461	22227	47112	76564
University of Lincoln and Humberside	ulh.ac.uk	5	718	268	1212	1405	2648
University of Ulster	ulst.ac.uk	120	9602	6915	23175	75717	110978
University of Manchester Institute of Science and Technology	umist.ac.uk	92	6186	3576	13475	32922	55886
University of North London	unl.ac.uk	48	7639	5490	17505	49112	75752
University of the West of England	uwe.ac.uk	39	11164	2397	19387	24725	54133
University of Wales Institute at Cardiff	uwic.ac.uk	21	1521	598	2599	3971	7176
University of Warwick	warwick.ac.uk	79	13030	5037	27860	47159	96206
University of Wolverhampton	wlv.ac.uk	44	16044	5226	29149	53085	90709
University of Westminster	wmin.ac.uk	38	6051	2025	10360	21174	32825
University College Worcester	worc.ac.uk	14	3087	1033	5641	7184	15439
University of York	york.ac.uk	57	17592	5808	35928	67257	123881

Appendix 2: Staff Number and RAE Rating Data for 111 UK Universities

University Name	Domain Name	Staff Numbers	RAE Rating
University of Aberdeen	abdn.ac.uk	684.7	4
University of Wales, Aberystwyth	aber.ac.uk	367.4	4
Anglia Polytechnic University	anglia.ac.uk	722.2	0.5
Aston University	aston.ac.uk	265.9	3.8
University of Wales Bangor	bangor.ac.uk	357.4	3.9
University of Bath	bath.ac.uk	480.3	5.2
Bath Spa University College	bathspa.ac.uk	164	1.5
University of Birmingham	bham.ac.uk	1381.9	4.3
University of Bournemouth	bournemouth.ac.uk	387.3	0.7
University of Bradford	brad.ac.uk	432.5	3.3
University of Bristol	bris.ac.uk	1169.7	5.2
Oxford Brookes University	brookes.ac.uk	621.5	1.6
Brunel University	brunel.ac.uk	550.5	3.1
University of Brighton	bton.ac.uk	605.9	1.7
University of Buckingham	buckingham.ac.uk	339.3	0.8
University of Cambridge	cam.ac.uk	1892.5	6.5
Canterbury Christ Church University College	cant.ac.uk	265.6	1.1
Cardiff University	cf.ac.uk	874.9	5
Chichester College	chichester.ac.uk	155.1	1.7
City University, London	city.ac.uk	347.2	3.7
Cranfield University	cranfield.ac.uk	625.8	1
University of Coventry	coventry.ac.uk	443.3	2.6
University of Derby	derby.ac.uk	445	0.9
De Montfort University	dmu.ac.uk	1188.5	1.4
University of Dundee	dundee.ac.uk	595.1	4.2
University of Durham	dur.ac.uk	696.4	5.2
University of Edinburgh	ed.ac.uk	1623.9	5

University of Essex	essex.ac.uk	398.2	4.8
University of Exeter	ex.ac.uk	547.7	4.7
Glasgow Caledonian University	gcal.ac.uk	616.7	1.2
University of Glasgow	gla.ac.uk	1386.9	4.2
University of Glamorgan	glam.ac.uk	458.9	1.3
University of Gloucestershire	glos.ac.uk	125.9	1.7
Goldsmiths College, University of London	goldsmiths.ac.uk	280.7	4.9
University of Greenwich	gre.ac.uk	767.6	1.2
Harper-Adams Agricultural College	harper-adams.ac.uk	100	1.1
University of Hertfordshire	herts.ac.uk	692.2	1.4
University of Huddersfield	hud.ac.uk	578.7	0.9
University of Hull	hull.ac.uk	535.3	3.2
Heriot-Watt University	hw.ac.uk	398.9	4.2
Imperial College, University of London	ic.ac.uk	1340.4	5.8
King’s College London	kcl.ac.uk	1448.6	4.7
University of Keele	keele.ac.uk	444.3	4
Kingston University	king.ac.uk	532.8	1.2
University of Wales, Lampeter	lamp.ac.uk	70	4.4
University of Lancaster	lancs.ac.uk	518.2	5.4
University of Loughborough	lboro.ac.uk	588.4	4.3
University of Leicester	le.ac.uk	788.1	4.5
University of Leeds	leeds.ac.uk	1472.7	4.5
London Guildhall University	lgu.ac.uk	418.7	0.8
University of Liverpool	liv.ac.uk	972.2	4.6
Liverpool John Moores University	livjm.ac.uk	820.1	1.2
Leeds Metropolitan University	lmu.ac.uk	692.9	0.8
London School of Economics	lse.ac.uk	444.1	6.3
University of Luton	luton.ac.uk	441.7	0.6
University of Manchester	man.ac.uk	1457.1	5

University of Middlesex	mdx.ac.uk	774.4	1.3
Manchester Metropolitan University	mmu.ac.uk	1186.4	1.2
Napier University	napier.ac.uk	495.5	0.8
University of Newcastle	ncl.ac.uk	1026.9	4.4
University of Wales College Newport	newport.ac.uk	160.3	0.5
Northampton University College	northampton.ac.uk	381.9	1
University of Northumbria	northumbria.ac.uk	824	1.1
University of Nottingham	nott.ac.uk	1127.2	5
Nottingham Trent University	ntu.ac.uk	886.4	1.4
Open University	open.ac.uk	986.6	2.5
University of Oxford	ox.ac.uk	2132.8	6.2
University of Paisley	paisley.ac.uk	414.3	0.7
University of Plymouth	plym.ac.uk	660	1.6
University of Portsmouth	port.ac.uk	708.2	1.7
Queen Margaret University College	qmced.ac.uk	193.3	1.5
Queen Mary, University of London	qmw.ac.uk	753.3	4.7
Queen’s University Belfast	qub.ac.uk	915.9	4.3
University of Reading	rdg.ac.uk	764.9	4.9
The Robert Gordon University	rgu.ac.uk	469.7	0.8
Royal Holloway, University of London	rhul.ac.uk	383.2	5.2
University of Salford	salford.ac.uk	702	2.1
South Bank University	sbu.ac.uk	595.3	1.3
University of Sheffield	shef.ac.uk	1260.7	4.5
Sheffield Hallam University	shu.ac.uk	959.7	1.1
School of Oriental and African Studies	soas.ac.uk	235.4	5.3
University of Southampton	soton.ac.uk	1039.3	5.4
University of St Andrews	st-and.ac.uk	401.5	5.3
University of Staffordshire	staffs.ac.uk	548.6	1.1
University of Stirling	stir.ac.uk	476.9	3.9

University of Strathclyde	strath.ac.uk	898.6	3.6
University of Sunderland	sunderland.ac.uk	520.8	1.7
The Surrey Institute of Art and Design University College	surrart.ac.uk	100	1.7
University of Surrey	surrey.ac.uk	520.5	4.7
University of Sussex	susx.ac.uk	566.3	5.1
University of Wales Swansea	swan.ac.uk	443	4.4
University of Abertay, Dundee	tay.ac.uk	231	0.7
University of Teeside	tees.ac.uk	441.4	0.7
Thames Valley University	tvu.ac.uk	157.6	0.4
University of Central England	uce.ac.uk	680.9	0.8
University College London	ucl.ac.uk	1985.3	5.4
University of Central Lancashire	uclan.ac.uk	796.6	0.9
University of East Anglia	uea.ac.uk	473	5
University of East London	uel.ac.uk	523.5	1.4
University of Kent at Canterbury	ukc.ac.uk	439.4	4
University of Lincoln and Humberside	ulh.ac.uk	349.7	0.7
University of Ulster	ulst.ac.uk	951.5	2.4
University of Manchester Institute of Science and Technology	umist.ac.uk	484.9	5.2
University of North London	unl.ac.uk	510.8	1.1
University of the West of England	uwe.ac.uk	810.2	1.6
University of Wales Institute at Cardiff	uwic.ac.uk	337.7	0.8
University of Warwick	warwick.ac.uk	786.6	5.6
University of Wolverhampton	wlv.ac.uk	744.2	0.6
University of Westminster	wmin.ac.uk	689.2	1.2
University College Worcester	worc.ac.uk	225.3	0.8
University of York	york.ac.uk	593.2	5.5

Appendix 3: Inlink and Outlink Page, Domain, Directory and Site ADM Data for New Zealand Universities 2000 – 2006

Jul 2000

Domain Name	Static Pages	Domains	Page inlinks	Directory inlinks	Domain inlinks	Site inlinks	Page outlinks	Directory outlinks	Domain outlinks	Site outlinks
aut.ac.nz	14253	1	5	5	2	1	0	0	0	0
lincoln.ac.nz	16766	1	51	41	13	1	0	0	0	0
massey.ac.nz	28463	1	111	95	39	1	0	0	0	0
auckland.ac.nz	36771	38	0	0	0	0	720	586	251	7
otago.ac.nz	47345	1	98	87	48	1	0	0	0	0
waikato.ac.nz	47193	1	192	151	57	1	0	0	0	0
canterbury.ac.nz	56012	1	125	96	52	1	0	0	0	0
vuw.ac.nz	180090	1	138	111	40	1	0	0	0	0

Feb 2002

Domain Name	Static Pages	Domains	Page inlinks	Directory inlinks	Domain inlinks	Site inlinks	Page outlinks	Directory outlinks	Domain outlinks	Site outlinks
aut.ac.nz	8843	3	57	48	31	7	51	40	25	7
lincoln.ac.nz	3140	1	317	216	78	8	48	39	23	6
massey.ac.nz	27172	21	490	396	171	8	232	177	98	7
auckland.ac.nz	67023	33	689	446	209	7	1419	1174	486	14
otago.ac.nz	73044	28	429	360	187	8	425	350	201	7
waikato.ac.nz	40118	15	671	512	211	8	465	346	145	7
canterbury.ac.nz	59757	39	498	405	219	8	490	392	213	7
vuw.ac.nz	36045	6	610	503	181	8	631	368	96	7

Dec 2003

Domain Name	Static Pages	Domains	Page inlinks	Directory inlinks	Domain inlinks	Site inlinks	Page outlinks	Directory outlinks	Domain outlinks	Site outlinks
aut.ac.nz	8575	3	144	122	65	13	135	121	78	14
lincoln.ac.nz	3590	1	508	357	142	15	92	71	46	12
massey.ac.nz	25734	29	894	730	348	15	480	373	241	14
auckland.ac.nz	69720	36	1700	1082	468	14	2410	1886	756	21
otago.ac.nz	43255	30	858	727	355	15	830	695	411	14
waikato.ac.nz	35594	13	1228	940	378	15	914	701	280	14
canterbury.ac.nz	29157	39	866	725	415	15	907	736	410	14
vuw.ac.nz	79241	11	1249	943	344	15	1679	1043	293	14

Jan 2003

Domain Name	Static Pages	Domains	Page inlinks	Directory inlinks	Domain inlinks	Site inlinks	Page outlinks	Directory outlinks	Domain outlinks	Site outlinks
aut.ac.nz	7061	3	200	174	97	19	208	189	126	21
lincoln.ac.nz	3530	1	789	526	212	22	138	108	70	18
massey.ac.nz	26294	21	1266	1023	492	22	754	589	363	21
auckland.ac.nz	62295	27	2458	1595	697	21	3238	2518	1004	28
otago.ac.nz	41295	28	1208	1021	501	22	1283	1090	641	21
waikato.ac.nz	39947	12	1705	1319	543	22	1339	1033	401	21
canterbury.ac.nz	29128	38	1286	1074	599	22	1406	1103	609	21
vuw.ac.nz	38478	7	1833	1372	491	22	2379	1474	418	21

Jan 2005

Domain Name	Static Pages	Domains	Page inlinks	Directory inlinks	Domain inlinks	Site inlinks	Page outlinks	Directory outlinks	Domain outlinks	Site outlinks
aut.ac.nz	12055	8	83	72	34	6	89	83	64	7
lincoln.ac.nz	3839	1	183	145	68	7	31	23	19	6
massey.ac.nz	26591	34	367	312	189	7	419	302	166	7
auckland.ac.nz	61447	39	854	620	270	7	954	737	325	7
otago.ac.nz	38564	34	391	323	168	7	416	342	211	7
waikato.ac.nz	45026	19	476	381	165	7	491	393	153	7
canterbury.ac.nz	29388	42	385	319	185	7	415	341	193	7
vuw.ac.nz	30047	4	636	443	167	7	560	394	115	7

Jan 2006

Domain Name	Static Pages	Domains	Page inlinks	Directory inlinks	Domain inlinks	Site inlinks	Page outlinks	Directory outlinks	Domain outlinks	Site outlinks
aut.ac.nz	14823	5	97	93	44	6	144	97	72	7
lincoln.ac.nz	703	1	166	147	62	7	27	15	10	6
massey.ac.nz	29663	39	365	306	192	7	514	365	195	7
auckland.ac.nz	66060	40	1021	684	300	7	977	746	314	7
otago.ac.nz	50039	36	449	364	184	7	439	355	206	7
waikato.ac.nz	48516	18	494	374	160	7	496	380	151	7
canterbury.ac.nz	28243	42	415	322	178	7	479	386	206	7
vuw.ac.nz	35437	5	682	476	159	7	613	422	125	7

Appendix 4: Inlink and Outlink Page, Domain, Directory and Site ADM Data for Australian Universities 2000 – 2006

Jul 2000

Domain Name	Static Pages	Domains	Page inlinks	Directory inlinks	Domain inlinks	Site inlinks	Page outlinks	Directory outlinks	Domain outlinks	Site outlinks
acu.edu.au	8873	1	248	198	98	32	89	83	64	7
adelaide.edu.au	65268	1	2393	1265	664	36	31	23	19	6
adfa.oz.au	1121	1	960	796	424	36	419	302	166	7
amc.edu.au	681	1	879	639	283	18	954	737	325	7
anu.edu.au	67033	74	4371	3224	1505	36	5280	4371	2318	85
ballarat.edu.au	6161	1	636	510	244	34	491	393	153	7
bond.edu.au	6658	1	633	524	286	36	415	341	193	7
canberra.edu.au	26534	1	1493	1116	583	36	560	394	115	7
cowan.edu.au	19190	1	658	423	224	30	0	0	0	0
cqu.edu.au	32928	20	405	332	212	27	821	652	448	37
csu.edu.au	29718	16	1657	1183	427	29	2738	1493	686	38
curtin.edu.au	11724	38	845	707	382	29	1121	933	541	38
deakin.edu.au	3316	9	746	590	316	29	293	228	155	32
flinders.edu.au	34727	29	896	730	429	28	1586	746	494	36
gu.edu.au	46149	21	1106	844	361	29	1336	983	339	38
jcu.edu.au	632	6	601	462	251	29	48	46	38	20
latrobe.edu.au	13160	32	937	796	426	29	961	842	556	37
monash.edu.au	56086	50	2359	1837	978	29	1737	1414	897	37
mq.edu.au	333	8	1294	1083	711	30	25	24	24	10
murdoch.edu.au	1	1	1165	901	527	30	0	0	0	0
newcastle.edu.au	29537	25	918	648	360	29	1325	1112	590	36
ntu.edu.au	27035	17	311	263	137	29	743	561	309	37
qut.edu.au	40275	38	1027	796	448	28	657	540	369	35
rmit.edu.au	38895	48	836	694	474	29	2355	1367	763	38
scu.edu.au	9797	12	546	469	200	29	1520	1148	461	38

swin.edu.au	29413	16	635	455	219	29	4148	673	379	37
une.edu.au	27883	15	582	454	225	28	957	835	394	35
unimelb.edu.au	56530	74	7181	2096	1295	29	667	587	448	37
unisa.edu.au	37266	29	690	579	361	28	1682	1320	600	38
unsw.edu.au	50522	92	1690	1412	863	30	1819	1564	949	37
uow.edu.au	30860	18	634	492	277	29	1715	1200	513	37
uq.edu.au	48040	48	1685	1315	715	29	1376	1258	692	38
usc.edu.au	582	1	23	21	18	13	0	0	0	0
usq.edu.au	21366	2	366	312	153	28	545	476	235	38
usyd.edu.au	50425	60	2223	1823	1004	30	3565	2329	1209	37
uts.edu.au	57517	42	996	778	461	30	1485	1020	650	38
uwa.edu.au	55658	70	1044	850	558	30	1438	1028	740	39
uws.edu.au	31766	23	708	551	356	30	842	722	457	39
vu.edu.au	12933	20	84	60	43	20	879	751	469	39
info.utas.edu.au	46747	1	0	0	0	0	707	431	209	39

Jan 2002

Domain Name	Static Pages	Domains	Page inlinks	Directory inlinks	Domain inlinks	Site inlinks	Page outlinks	Directory outlinks	Domain outlinks	Site outlinks
acu.edu.au	11228	7	305	214	121	33	535	442	231	36
adelaide.edu.au	127323	51	2606	1850	983	36	5138	2182	1151	36
anu.edu.au	15366	73	8933	6309	2668	37	1552	1348	793	37
ballarat.edu.au	5040	1	307	248	129	35	61	57	39	19
bond.edu.au	12666	7	408	322	166	33	363	314	172	37
canberra.edu.au	4956	16	1306	994	606	35	392	318	225	34
cowan.edu.au	17669	9	2463	1052	457	35	768	580	249	36
cqu.edu.au	52164	31	741	628	417	35	1267	972	554	37
csu.edu.au	99842	24	3423	2285	769	37	5942	2732	1255	37
curtin.edu.au	75177	59	2323	1510	742	36	2595	1660	812	37
deakin.edu.au	43080	14	1543	1122	596	36	1177	802	446	37

flinders.edu.au	28795	32	1684	1311	837	36	1127	1002	615	37
gu.edu.au	68565	21	2154	1602	609	36	1296	1015	470	37
jcu.edu.au	98248	22	1115	825	453	36	8629	2093	1017	37
latrobe.edu.au	65240	41	1852	1462	767	36	5207	2386	1184	37
monash.edu.au	177316	76	4912	3500	1822	37	3662	2846	1487	36
mq.edu.au	77642	77	2462	2020	1137	36	3539	2046	1203	37
murdoch.edu.au	52169	61	2945	1854	935	37	3209	2252	1025	37
newcastle.edu.au	74037	26	1715	1223	589	36	1884	1515	670	37
ntu.edu.au	29037	18	497	425	243	36	1260	802	384	37
qut.edu.au	82509	59	2618	1604	887	37	1435	1042	671	36
rmit.edu.au	75666	67	2002	1507	902	36	5972	1415	821	37
scu.edu.au	16603	8	1113	920	356	35	1834	1326	468	37
swin.edu.au	24887	15	1294	887	426	35	5242	560	319	34
une.edu.au	23036	17	1173	880	389	35	1161	982	446	36
unimelb.edu.au	201916	172	27600	3797	2235	37	4092	3219	1890	37
unisa.edu.au	88488	33	1667	1218	674	36	2358	1704	868	37
unsw.edu.au	251367	133	4184	2905	1635	36	3811	3048	1692	37
uow.edu.au	74210	23	1357	989	521	35	1182	1004	512	36
uq.edu.au	179174	98	3637	2648	1323	37	5417	4160	2055	37
usc.edu.au	1280	2	82	67	52	29	583	55	33	15
usq.edu.au	48844	3	807	626	265	36	6760	797	343	37
usyd.edu.au	200002	97	6101	3658	1842	37	9532	5088	2237	37
uts.edu.au	63066	61	3192	1432	852	36	3450	2048	1103	37
uwa.edu.au	121958	110	2707	1998	1211	36	2966	2102	1402	37
uws.edu.au	35747	15	1118	859	575	35	883	731	392	36
vu.edu.au	18428	17	3022	329	189	35	858	748	466	37
info.utas.edu.au	83734	1	704717	12207	952	36	704946	11894	632	38

Mar 2003

Domain Name	Static Pages	Domains	Page inlinks	Directory inlinks	Domain inlinks	Site inlinks	Page outlinks	Directory outlinks	Domain outlinks	Site outlinks
acu.edu.au	6371	5	252	190	118	32	543	462	204	37
adelaide.edu.au	104898	53	2374	1724	876	37	7640	1757	981	36
anu.edu.au	25861	76	7288	5087	2295	37	2440	2026	1051	37
ballarat.edu.au	5304	1	272	221	118	33	76	65	48	24
bond.edu.au	12831	10	340	278	143	36	235	224	96	28
canberra.edu.au	20517	23	1042	833	510	36	1105	761	334	35
cowan.edu.au	9867	5	1106	811	407	35	368	328	155	34
cqu.edu.au	61371	20	589	487	329	37	1161	879	457	37
csu.edu.au	80015	19	2694	1926	650	37	3893	2136	972	37
curtin.edu.au	66220	53	1855	1326	690	36	1993	1312	755	36
deakin.edu.au	36321	12	1293	976	524	37	1209	915	433	36
flinders.edu.au	26829	29	1737	1352	739	37	992	891	529	37
gu.edu.au	90848	15	1721	1341	512	36	828	633	350	37
jcu.edu.au	92730	21	969	713	398	36	9897	1894	869	37
latrobe.edu.au	76089	20	1570	1270	652	36	4109	2271	1133	37
monash.edu.au	160416	64	4116	3009	1594	37	3397	2586	1370	37
mq.edu.au	50166	71	2286	1881	950	37	1886	1530	916	37
murdoch.edu.au	46339	64	2304	1497	772	37	1766	1470	879	37
newcastle.edu.au	45907	19	1934	1209	501	36	2618	949	435	37
ntu.edu.au	16234	13	527	427	210	35	862	447	199	34
qut.edu.au	59110	55	1529	1279	760	37	1561	1144	598	36
rmit.edu.au	122381	48	1762	1331	814	37	5529	1085	644	35
scu.edu.au	10584	10	811	689	301	36	1732	1296	520	37
swin.edu.au	42732	18	936	726	350	35	6646	700	413	36
une.edu.au	27558	15	950	736	367	36	1247	1059	463	37
unimelb.edu.au	188372	168	23494	3516	2073	37	4014	3019	1786	36
unisa.edu.au	29890	22	1450	1053	582	37	1991	1458	660	37

unsw.edu.au	219317	136	4034	2656	1521	37	3318	2583	1516	37
uow.edu.au	67988	28	1143	888	494	37	2096	1426	600	37
uq.edu.au	94226	85	3452	2451	1279	37	2887	2270	1258	37
usc.edu.au	1109	2	77	66	47	27	129	123	67	37
usq.edu.au	86217	4	636	518	227	37	1184	1027	436	37
usyd.edu.au	152870	100	4982	3262	1677	37	6956	4009	1955	37
utas.edu.au	46082	37	1333	1003	664	35	1457	1099	612	37
uts.edu.au	43122	56	2945	1238	749	36	3065	1784	952	37
uwa.edu.au	178139	138	2319	1674	1086	37	2550	1896	1220	37
uws.edu.au	26428	16	869	698	458	36	645	545	350	35
vu.edu.au	26575	20	5836	393	215	34	802	676	436	36

Feb 2004

Domain Name	Static Pages	Domains	Page inlinks	Directory inlinks	Domain inlinks	Site inlinks	Page outlinks	Directory outlinks	Domain outlinks	Site outlinks
acu.edu.au	4935	7	564	392	242	66	779	631	334	70
adelaide.edu.au	45686	49	4647	3416	1775	74	9794	3416	1933	72
anu.edu.au	141379	114	13948	9546	4461	74	7484	5890	3081	74
ballarat.edu.au	15775	3	582	440	239	67	241	214	164	50
bond.edu.au	19283	10	659	546	280	72	470	434	209	59
canberra.edu.au	19312	19	2213	1707	1047	72	2037	1354	652	70
cdu.edu.au	15553	10	1684	1261	630	69	1257	757	331	67
cowan.edu.au	27347	19	2094	1312	761	73	2283	1812	900	74
cqu.edu.au	31768	33	3362	2418	990	74	5078	2958	1483	74
csu.edu.au	66114	16	4449	3176	1297	73	5528	3315	1673	73
curtin.edu.au	51145	56	3130	2339	1280	74	4082	2256	1121	72
deakin.edu.au	43802	13	2940	2261	1244	74	2061	1742	965	74
flinders.edu.au	20960	19	3371	2609	1266	73	1883	1571	846	74
gu.edu.au	72666	16	3031	2116	911	72	10877	2620	1232	74
jcu.edu.au	65418	18	2439	2000	1057	73	7366	4221	1948	74

latrobe.edu.au	47568	20	5664	4202	2203	73	6810	4535	2331	74
monash.edu.au	175792	68	5992	4662	2514	74	5752	4359	2407	74
mq.edu.au	78066	72	4512	3250	1744	74	4061	3307	1968	74
murdoch.edu.au	37836	65	4506	2700	1279	73	4192	2166	1195	74
newcastle.edu.au	38862	17	1767	1419	691	70	2590	1418	651	71
qut.edu.au	39684	64	3169	2548	1545	74	2473	1858	1024	71
rmit.edu.au	90569	50	3435	2591	1570	74	10850	2121	1281	71
scu.edu.au	14882	10	1642	1373	603	72	2984	2310	1005	74
swin.edu.au	44842	16	1817	1404	705	71	7672	1416	856	73
une.edu.au	24023	17	1947	1497	753	71	2379	2034	927	74
unimelb.edu.au	194198	167	32866	7063	4183	74	7441	5594	3446	72
unisa.edu.au	33583	20	2940	2181	1204	74	3624	2614	1273	74
unsw.edu.au	163822	149	8526	5463	3153	74	6428	4925	2864	74
uow.edu.au	32023	26	2264	1792	1007	74	3252	2415	1141	74
uq.edu.au	119158	85	7415	5118	2626	74	6709	4769	2475	74
usc.edu.au	1148	2	172	147	104	56	262	250	136	74
usq.edu.au	82677	4	1284	1015	445	74	2607	2112	864	74
usyd.edu.au	138025	99	9065	6395	3339	74	13027	7516	3779	74
utas.edu.au	52718	32	2627	2001	1319	71	2865	2222	1254	74
uts.edu.au	44775	56	6017	2492	1533	72	4821	3053	1769	74
uwa.edu.au	150500	142	5051	3542	2230	74	4677	3662	2389	74
uws.edu.au	22680	11	1769	1369	888	73	1266	1038	654	72
vu.edu.au	39365	22	6270	692	425	68	1868	1570	982	73

Mar 05

Domain Name	Static Pages	Domains	Page inlinks	Directory inlinks	Domain inlinks	Site inlinks	Page outlinks	Directory outlinks	Domain outlinks	Site outlinks
acu.edu.au	6193	7	870	591	367	101	1048	817	465	103
adelaide.edu.au	59288	55	7246	5097	2647	110	11892	5052	2879	108
anu.edu.au	149986	119	20371	13868	6577	110	13398	9872	5161	111

ballarat.edu.au	18341	4	903	680	369	99	445	393	285	79
bond.edu.au	17540	8	987	806	414	105	663	612	318	92
canberra.edu.au	21570	16	3201	2467	1541	107	2834	1843	921	105
cdu.edu.au	14334	10	2218	1665	852	102	2329	1283	540	103
cqu.edu.au	33098	36	3509	2086	1128	109	3447	2624	1432	111
csu.edu.au	70744	15	5630	4089	1545	110	9027	4913	2347	111
curtin.edu.au	69763	69	6639	4569	2061	109	7647	4455	2242	107
deakin.edu.au	53239	9	4366	3272	1769	110	4959	2932	1434	109
ecu.edu.au	29308	17	4574	3101	1653	108	2965	2471	1309	108
flinders.edu.au	26410	20	4896	3766	1974	109	3031	2579	1400	111
gu.edu.au	76352	15	4502	3262	1390	108	11812	3327	1585	111
jcu.edu.au	60450	18	3358	2743	1459	109	10893	6421	2861	111
latrobe.edu.au	40827	14	7250	5319	2756	109	10199	6501	3242	111
monash.edu.au	161421	72	9738	7264	3970	110	9829	7161	3864	111
mq.edu.au	68767	68	6632	4917	2708	110	6506	5205	2977	111
murdoch.edu.au	33541	67	6843	4201	2037	109	5662	3217	1917	111
newcastle.edu.au	34289	19	3143	2452	1156	105	4143	2175	984	108
qut.edu.au	25948	46	4924	3819	2328	110	3014	2297	1269	106
rmit.edu.au	56881	48	5398	3891	2285	109	15938	3175	1946	108
scu.edu.au	9374	8	2451	2038	891	107	4165	3244	1468	111
une.edu.au	28228	17	2949	2230	1130	106	3551	3070	1425	111
unimelb.edu.au	159104	173	42364	10711	6347	110	10751	8113	5150	109
unisa.edu.au	48473	20	4422	3264	1796	110	5420	3832	1914	111
unsw.edu.au	210431	145	13205	8234	4771	110	9815	7465	4362	111
uow.edu.au	50470	35	3707	2800	1552	110	5123	3604	1779	111
uq.edu.au	98904	88	11713	7734	3969	110	10509	7136	3735	111
usc.edu.au	1561	2	264	227	162	85	391	371	217	111
usq.edu.au	93550	5	1913	1484	657	110	4052	3181	1290	111
usyd.edu.au	151412	131	13110	9484	4971	110	19979	11260	5769	111

utas.edu.au	50889	32	3930	2988	1990	107	4206	3297	1885	111
uts.edu.au	51608	61	9322	3802	2340	107	7049	4573	2692	111
uwa.edu.au	157347	171	7821	5378	3414	110	7427	5803	3749	111
uws.edu.au	10844	10	2694	2040	1313	109	1781	1423	891	109
vu.edu.au	39747	17	6756	1010	650	103	2967	2340	1437	110

Apr 2006

Domain Name	Static Pages	Domains	Page inlinks	Directory inlinks	Domain inlinks	Site inlinks	Page outlinks	Directory outlinks	Domain outlinks	Site outlinks
acu.edu.au	4960	6	277	177	116	34	193	145	110	31
adelaide.edu.au	59575	54	2510	1575	783	37	2055	1638	860	36
anu.edu.au	289052	128	5276	3806	1929	37	5840	3991	1984	37
ballarat.edu.au	22024	3	271	194	109	34	271	240	149	31
bond.edu.au	11277	6	282	227	116	31	169	148	89	30
canberra.edu.au	12762	14	952	712	452	36	533	458	278	34
cdu.edu.au	15679	9	576	409	209	35	1111	616	242	35
ecu.edu.au	36925	19	1560	772	360	36	1724	1232	443	35
cqu.edu.au	16068	28	1265	602	317	37	653	468	316	37
csu.edu.au	57052	13	2079	1342	468	37	1818	1194	593	37
curtin.edu.au	74659	86	2172	1301	751	37	2389	1200	604	35
deakin.edu.au	40118	7	1191	910	450	37	883	641	275	37
flinders.edu.au	24330	21	1501	1107	632	36	1754	1317	666	37
gu.edu.au	70179	17	1078	861	384	36	1063	809	404	37
jcu.edu.au	84770	21	864	697	345	37	4762	2933	995	37
latrobe.edu.au	40713	16	1443	1025	474	37	2838	1687	832	37
monash.edu.au	160583	62	3011	2336	1262	37	3624	2576	1363	37
mq.edu.au	54979	77	2127	1543	858	37	2243	1737	1001	37
murdoch.edu.au	41440	64	1999	1271	627	37	2443	1328	746	37
newcastle.edu.au	23593	11	1117	886	423	36	2503	745	311	37
qut.edu.au	38764	60	1567	1108	675	37	826	662	387	37

rmit.edu.au	32237	36	1666	1133	618	36	770	596	404	37
scu.edu.au	5455	6	652	537	248	36	334	265	183	37
swinburne.edu.au	17451	18	830	626	310	36	390	321	238	37
une.edu.au	21687	16	969	696	352	36	633	542	311	37
unimelb.edu.au	162922	171	6110	3316	1936	37	3112	2478	1692	37
unisa.edu.au	51600	17	1418	1045	530	37	1641	1068	543	37
unsw.edu.au	120840	110	4335	2562	1507	37	1881	1561	1028	36
uow.edu.au	34999	32	1456	1012	531	37	1616	947	512	36
uq.edu.au	102735	67	4650	2752	1292	37	2781	1737	919	37
usc.edu.au	1879	1	81	73	48	27	102	95	64	37
usq.edu.au	95879	5	533	404	192	37	983	841	387	37
usyd.edu.au	178207	135	3849	2871	1544	37	8103	3952	2029	37
utas.edu.au	54142	38	1343	954	628	36	912	719	466	35
uts.edu.au	46076	64	3256	1302	763	37	1758	1246	806	37
uwa.edu.au	138260	162	3170	1885	1124	37	2113	1540	1025	35
uws.edu.au	7802	5	744	572	363	37	331	233	150	36
vu.edu.au	34850	27	443	298	209	35	1468	993	530	37

Appendix 5: Inlink and Outlink Page, Domain, Directory and Site ADM Data for UK Universities 2000 – 2005

Jul 2000

Domain Name	Static Pages	Domains	Page inlinks	Directory inlinks	Domain inlinks	Site inlinks	Page outlinks	Directory outlinks	Domain outlinks	Site outlinks
abdn.ac.uk	7253	9	1282	1056	428	84	1262	857	328	82
aber.ac.uk	31886	9	1152	859	289	85	2829	1960	572	94
abertay-dundee.ac.uk	219	1	539	261	86	43	406	143	11	4
anglia.ac.uk	5620	10	647	409	195	67	256	243	151	63
aston.ac.uk	25760	16	419	356	215	70	930	815	539	86
bangor.ac.uk	15472	17	702	561	281	76	730	596	332	70
bath.ac.uk	8246	3	1669	1282	435	81	827	742	300	75
bbk.ac.uk	26661	18	1075	749	369	80	2381	1772	573	82
brad.ac.uk	8073	12	638	567	247	77	964	841	391	77
bris.ac.uk	3977	15	2979	2366	1057	93	653	524	257	78
brookes.ac.uk	2877	2	501	401	167	78	211	166	88	55
brunel.ac.uk	7007	3	862	692	360	85	525	457	254	86
bton.ac.uk	4603	10	674	505	262	72	475	344	188	63
buckingham.ac.uk	3404	2	18	16	12	12	30	30	26	12
cf.ac.uk	30484	13	3139	1586	520	85	1951	1725	637	91
city.ac.uk	9109	16	795	668	366	76	760	643	385	79
courtauld.ac.uk	156	1	49	42	21	17	11	7	6	5
coventry.ac.uk	10469	13	239	203	134	60	763	677	357	97
cranfield.ac.uk	10041	6	540	465	217	80	639	515	278	84
derby.ac.uk	10651	7	331	289	174	67	365	237	147	65
dmu.ac.uk	10968	33	945	750	370	82	1138	1017	550	80
dundee.ac.uk	12787	17	1015	870	385	83	813	723	377	79
dur.ac.uk	13573	14	1607	1221	433	81	3206	2132	433	92
ed.ac.uk	15645	84	6692	4967	2063	92	1550	1377	885	89

essex.ac.uk	16822	26	1562	1275	633	80	1214	1066	583	92
ex.ac.uk	7915	6	1501	1252	428	92	1086	935	351	102
goldsmiths.ac.uk	1288	1	445	354	165	70	140	117	77	59
gre.ac.uk	7605	8	270	231	118	66	628	574	217	76
herts.ac.uk	12612	14	644	521	263	74	919	726	376	81
heythrop.ac.uk	205	1	21	21	13	12	40	38	30	27
hud.ac.uk	13425	6	387	280	131	67	1255	1080	245	81
hw.ac.uk	6444	13	2504	2040	778	91	2734	1831	704	91
icr.ac.uk	30287	3	93	81	58	37	110	75	35	23
ioe.ac.uk	3065	3	244	208	86	52	292	237	133	64
ion.ac.uk	1	1	5	4	3	2	0	0	0	0
kcl.ac.uk	32971	17	1515	1104	419	89	3283	1882	681	93
keele.ac.uk	19101	7	833	615	264	83	1333	1147	418	82
kingston.ac.uk	3075	4	282	236	130	64	451	373	193	71
lamp.ac.uk	36751	3	278	207	82	47	1675	1342	160	61
lboro.ac.uk	8499	11	891	742	337	78	835	636	296	78
lbs.ac.uk	3449	1	286	191	42	21	39	38	20	15
le.ac.uk	13554	14	1465	1163	448	84	2058	1706	594	94
leeds.ac.uk	17033	39	4193	3447	1106	90	2596	1992	788	91
lgu.ac.uk	8604	1	542	407	149	79	514	461	202	80
liv.ac.uk	6722	9	1825	1378	430	87	1512	1181	363	93
livjm.ac.uk	13879	10	455	375	191	69	893	810	427	89
lmu.ac.uk	6973	1	222	186	92	66	390	334	163	68
lon.ac.uk	3947	3	1368	422	202	65	341	250	65	29
lse.ac.uk	21703	14	766	532	285	75	866	462	221	65
lshtm.ac.uk	1344	1	283	111	43	30	114	90	56	40
luton.ac.uk	1461	2	138	124	64	46	91	80	51	40
mbs.ac.uk	167	1	48	43	28	20	1	1	1	1
napier.ac.uk	10179	6	559	470	230	67	1172	892	313	75

ncl.ac.uk	313	1	2533	2100	802	86	3	3	3	3
newport.ac.uk	903	5	128	90	36	24	72	67	51	36
nott.ac.uk	10245	18	1972	1656	763	86	1404	1205	511	83
open.ac.uk	10182	33	1420	1216	736	84	1423	1204	629	88
paisley.ac.uk	3108	4	170	141	90	42	525	447	211	85
plymouth.ac.uk	5319	9	496	436	291	75	295	231	155	66
port.ac.uk	7709	30	453	409	265	74	1143	964	590	88
qmw.ac.uk	8563	29	1225	989	586	84	1098	831	444	82
qub.ac.uk	12437	13	1085	875	432	80	1498	1141	378	85
rcm.ac.uk	49	1	30	22	14	13	0	0	0	0
rcplondon.ac.uk	604	1	8278	3408	348	72	11	11	8	8
rdg.ac.uk	10900	25	1001	844	377	81	9959	4529	559	85
rgu.ac.uk	8456	6	350	296	148	70	1013	801	384	64
rhbnc.ac.uk	26862	29	516	439	277	73	3658	2266	981	91
salford.ac.uk	18818	12	527	445	198	75	899	693	348	71
sas.ac.uk	5244	3	777	577	193	71	624	308	185	74
sbu.ac.uk	16712	12	634	505	237	80	1283	1025	458	105
shef.ac.uk	2797	1	3386	2620	649	89	216	190	102	59
shu.ac.uk	12734	7	656	525	209	77	1423	979	315	91
soas.ac.uk	1866	3	251	183	77	49	188	119	63	40
soton.ac.uk	55912	28	5740	3518	1108	90	1632	1342	571	94
ssees.ac.uk	2148	1	101	66	38	27	47	35	22	17
st-and.ac.uk	13237	41	182741	14882	782	74	183317	15555	943	90
stir.ac.uk	17749	17	833	654	310	77	1146	965	421	84
stmarys-belfast.ac.uk	176	1	1094	164	10	8	1049	130	7	5
stran-ni.ac.uk	1223	1	4150	1009	17	13	4239	1103	49	29
strath.ac.uk	15539	39	1525	1228	604	86	1427	1095	616	84
sunderland.ac.uk	18973	15	492	404	217	70	1612	1142	525	87
surrey.ac.uk	9611	21	1414	1173	521	84	877	721	418	82

susx.ac.uk	10169	21	1268	1032	446	82	2106	1349	576	83
swan.ac.uk	20724	17	1017	769	283	78	2396	1771	493	96
tees.ac.uk	2038	7	201	184	118	51	182	166	133	77
tvu.ac.uk	7001	5	81	63	42	34	83	77	52	32
uce.ac.uk	3508	5	183	149	73	46	286	249	150	67
ucl.ac.uk	37541	63	3984	3029	1192	94	2502	1926	961	91
uclan.ac.uk	11356	4	242	206	107	65	764	626	277	89
uea.ac.uk	14557	15	1292	979	481	83	1391	1284	582	88
uel.ac.uk	7332	5	318	278	130	61	463	413	220	75
ukc.ac.uk	0	1	1234	1017	503	86	0	0	0	0
ulh.ac.uk	824	1	53	37	27	23	47	42	36	31
ulsop.ac.uk	92	2	12	11	7	7	9	9	8	8
ulst.ac.uk	34040	16	846	715	367	84	1427	1023	490	86
umds.ac.uk	2342	7	394	347	175	55	833	626	239	84
umist.ac.uk	8631	33	691	583	384	74	858	711	360	73
unl.ac.uk	3131	6	333	301	172	70	293	246	147	62
unn.ac.uk	8664	10	647	481	206	80	912	764	286	79
uwcm.ac.uk	14040	6	275	217	105	56	2201	799	183	69
uwe.ac.uk	17355	6	387	296	165	63	832	744	338	83
uwic.ac.uk	2341	7	113	72	43	22	197	128	93	47
wales.ac.uk	136	1	155	95	16	12	35	32	10	10
wlv.ac.uk	100629	1	1473	1218	494	94	0	0	0	0
wmin.ac.uk	23848	12	508	408	201	78	1404	1257	729	82
wye.ac.uk	263	1	43	42	30	24	56	49	26	21
york.ac.uk	46955	22	2109	1711	573	84	2356	1983	815	91

Jul 2001

Domain Name	Static Pages	Domains	Page inlinks	Directory inlinks	Domain inlinks	Site inlinks	Page outlinks	Directory outlinks	Domain outlinks	Site outlinks
abdn.ac.uk	67440	17	2774	2293	964	100	4023	3110	945	97
aber.ac.uk	51752	14	2123	1629	656	102	3507	2310	1024	100
anglia.ac.uk	22788	12	1405	804	470	90	1754	1086	640	93
aston.ac.uk	20472	17	1036	769	480	86	1299	982	622	94
bangor.ac.uk	15308	24	1434	1084	642	100	920	762	495	97
bath.ac.uk	45267	7	3623	2612	928	98	3623	2976	932	99
bathspa.ac.uk	1588	3	145	126	79	57	459	385	187	101
bham.ac.uk	307787	65	11517	7982	2090	103	10705	5780	2228	106
bournemouth.ac.uk	19781	12	742	568	309	93	585	485	313	87
brad.ac.uk	65201	24	1447	1200	569	99	2392	1850	906	96
bris.ac.uk	120311	79	7042	5394	2390	104	8329	6114	2878	102
brookes.ac.uk	18008	9	1179	932	395	95	1472	1166	436	99
brunel.ac.uk	70749	4	1972	1495	781	98	1808	1563	669	99
bton.ac.uk	45410	19	1398	1104	589	99	14367	1241	565	91
buckingham.ac.uk	1628	1	40	37	27	23	29	29	28	20
cam.ac.uk	323855	259	38262	13573	5994	105	20328	13478	5764	104
cant.ac.uk	14686	3	237	211	114	66	505	413	200	75
cf.ac.uk	54039	14	3613	2859	1290	101	3331	2836	994	101
chichester.ac.uk	239	1	16	16	11	9	58	57	55	53
city.ac.uk	63856	12	1864	1478	810	98	3800	1201	527	90
coventry.ac.uk	14541	17	532	432	327	88	2766	975	537	105
derby.ac.uk	17800	10	673	561	367	92	410	381	273	85
dmu.ac.uk	59070	43	2116	1482	767	98	2147	1723	951	94
dundee.ac.uk	33464	18	2336	1795	857	101	1864	1520	759	93
dur.ac.uk	69544	23	4498	3354	1143	99	5732	3954	1234	101
ed.ac.uk	182965	147	19778	11636	4585	104	9217	7442	3388	104
essex.ac.uk	51368	24	3332	2653	1338	100	2254	1868	975	98

ex.ac.uk	100519	14	3365	2783	1033	104	5093	3668	1356	106
gcal.ac.uk	9392	23	739	514	331	79	1447	962	565	86
gla.ac.uk	166392	102	9279	6406	2639	104	10359	7803	3344	105
glam.ac.uk	11487	6	412	367	242	84	689	569	319	79
goldsmiths.ac.uk	9692	13	778	597	343	87	579	475	319	94
gre.ac.uk	21516	9	612	481	248	88	1578	831	346	87
harper-adams.ac.uk	385	1	1397	245	32	26	1348	206	3	3
herts.ac.uk	36642	14	1174	972	532	97	2595	1673	615	95
hud.ac.uk	18352	11	805	613	285	88	1429	1235	365	90
hull.ac.uk	17281	14	2114	1564	593	101	1960	1526	648	94
hw.ac.uk	56491	31	5893	4369	1605	104	9717	6233	1910	101
ic.ac.uk	262906	107	24030	7913	3601	102	12320	7430	2089	93
kcl.ac.uk	76801	25	3149	2089	962	97	4446	3462	1266	97
keele.ac.uk	23996	8	1662	1278	549	102	1904	1437	586	95
king.ac.uk	18751	6	613	502	291	87	945	774	439	91
lamp.ac.uk	3424	5	430	334	168	71	330	265	149	58
lancs.ac.uk	199209	31	3603	2890	1242	102	4952	3336	1373	100
lboro.ac.uk	46710	26	1911	1571	699	98	3393	2593	1116	98
le.ac.uk	72840	14	3838	2923	1104	100	3794	3041	1087	101
leeds.ac.uk	135231	71	12752	10507	2651	105	10277	7249	2665	103
lgu.ac.uk	9723	1	1746	1390	489	102	1256	1059	450	178
liv.ac.uk	26634	11	3931	3032	1091	102	1267	1023	435	80
livjm.ac.uk	21539	11	901	708	436	96	965	836	440	95
lmu.ac.uk	11090	1	513	393	197	89	815	687	215	78
lse.ac.uk	24813	16	1940	1234	593	96	911	709	394	85
luton.ac.uk	4085	1	230	204	136	69	13	11	7	7
man.ac.uk	69461	94	10655	7441	3196	104	8296	5108	2439	102
mdx.ac.uk	50047	22	775	624	381	88	1930	1443	875	105
mmu.ac.uk	36025	28	1585	1234	664	99	3187	2278	1206	98
napier.ac.uk	89817	18	1700	1170	556	92	3857	2618	834	89

ncl.ac.uk	122764	62	5459	4356	1788	102	5385	3992	1792	98
newport.ac.uk	2625	11	126	93	60	41	255	191	139	60
northampton.ac.uk	4750	3	75	59	39	32	602	504	225	80
nott.ac.uk	100913	42	8855	4133	1815	101	4024	3160	1503	96
ntu.ac.uk	19719	26	1294	1020	662	99	1771	1283	732	101
open.ac.uk	51943	55	3529	2747	1668	101	2402	1870	957	96
ox.ac.uk	225384	229	24041	15687	6332	104	16449	10609	4648	103
paisley.ac.uk	27268	16	400	331	229	66	1081	737	433	91
plym.ac.uk	21148	25	1149	959	646	95	1595	1291	774	95
port.ac.uk	53212	27	945	818	571	91	3560	1916	998	96
qmced.ac.uk	7453	4	421	254	139	58	672	572	289	80
qmw.ac.uk	41498	31	3499	2614	1281	93	6213	2515	1069	90
qub.ac.uk	71076	35	2136	1733	943	98	3989	2589	1117	97
rdg.ac.uk	89765	33	9074	4093	1203	103	5484	3254	1109	95
rgu.ac.uk	8405	5	864	670	308	92	1093	881	435	84
rhul.ac.uk	37311	22	1319	1032	650	91	2398	1729	949	92
salford.ac.uk	15582	48	1396	914	463	92	804	641	426	79
sbu.ac.uk	35295	23	1641	1131	519	101	11365	5657	1442	106
shef.ac.uk	97261	20	7354	5777	1454	103	7770	5843	1456	104
shu.ac.uk	45925	16	1560	1197	457	98	2547	1833	642	101
soas.ac.uk	2974	2	392	335	177	70	266	215	92	49
soton.ac.uk	111568	102	14663	7887	2696	103	9778	7111	2493	106
staffs.ac.uk	140602	8	900	726	375	94	1309	1061	545	98
st-and.ac.uk	63439	51	1167174	142203	1620	95	1166610	142553	1609	98
stir.ac.uk	24026	26	1748	1400	685	100	1781	1386	608	92
strath.ac.uk	95724	80	4278	3137	1369	101	4440	3090	1602	95
sunderland.ac.uk	64874	15	1061	833	483	94	2690	1845	802	95
surrart.ac.uk	165	0	75	63	43	32	0	0	0	0
surrey.ac.uk	124011	25	3337	2636	1205	99	2757	2201	1067	94
susx.ac.uk	78479	25	3470	2667	1057	98	4747	2928	1051	91

swan.ac.uk	31811	27	1740	1329	593	93	2571	1852	731	100
tay.ac.uk	1757	7	257	209	149	66	412	112	73	44
tees.ac.uk	15954	14	544	444	251	81	721	533	364	96
tvu.ac.uk	13313	7	149	128	104	60	142	114	79	45
uce.ac.uk	16186	14	351	292	179	81	443	374	243	77
ucl.ac.uk	183442	103	49142	8933	2749	104	8646	6475	2843	97
uclan.ac.uk	27741	3	514	428	232	91	1056	830	359	92
uea.ac.uk	37123	17	2703	2223	1081	104	2189	1636	726	99
uel.ac.uk	16997	6	565	484	280	83	822	685	430	97
ukc.ac.uk	45750	11	2718	2140	1033	102	2580	1980	795	105
ulh.ac.uk	8552	2	139	103	74	45	161	133	96	54
ulst.ac.uk	97950	31	1849	1353	725	102	2234	1512	766	99
umist.ac.uk	55528	49	2855	1830	1052	96	2666	1534	728	81
unl.ac.uk	20278	5	0	0	0	0	0	0	0	0
unn.ac.uk	9615	11	1738	1036	480	101	1076	866	340	85
uwe.ac.uk	19198	9	1004	768	415	90	852	701	293	92
uwic.ac.uk	3317	9	175	114	83	48	155	133	94	46
warwick.ac.uk	49241	21	6238	4920	1774	104	52324	3777	1658	105
wlv.ac.uk	57350	10	2764	2147	989	103	3848	2248	1066	108
wmin.ac.uk	18213	11	1124	757	434	96	685	600	384	82
worc.ac.uk	5347	2	125	107	78	50	563	490	164	79
york.ac.uk	110644	26	5994	4406	1492	101	5280	3823	1434	102

Jul 02

Domain Name	Static Pages	Domains	Page inlinks	Directory inlinks	Domain inlinks	Site inlinks	Page outlinks	Directory outlinks	Domain outlinks	Site outlinks
abdn.ac.uk	103874	21	2567	2063	892	99	4528	3480	1015	102
aber.ac.uk	74963	14	1940	1504	648	101	3270	2365	1031	104
anglia.ac.uk	14716	11	1233	688	406	84	905	765	428	93
aston.ac.uk	12857	22	836	635	416	86	1305	1090	680	101
bangor.ac.uk	17370	22	1369	994	579	98	1042	871	554	99
bath.ac.uk	51904	11	3010	2206	820	96	4196	3103	1114	107
bathspa.ac.uk	2014	4	176	141	90	56	471	415	203	103
bham.ac.uk	260257	96	12573	8568	2047	101	8640	5529	2440	109
bournemouth.ac.uk	23030	13	677	494	300	91	584	469	328	85
brad.ac.uk	29697	16	1366	980	495	95	2439	1915	952	96
bris.ac.uk	113803	71	6091	4654	2224	105	7676	5590	2572	106
brookes.ac.uk	23399	8	1124	907	395	98	1829	1269	407	92
brunel.ac.uk	50706	4	1681	1317	694	96	1840	1566	654	100
bton.ac.uk	49346	21	1281	987	535	94	14843	1399	638	95
buckingham.ac.uk	1430	1	45	42	33	29	18	18	18	14
cam.ac.uk	206656	244	16912	11561	5484	102	20091	12330	5061	106
cant.ac.uk	17067	20	242	204	122	70	821	646	408	81
cf.ac.uk	44351	15	3396	2649	1197	100	2725	2322	879	106
chichester.ac.uk	211	1	15	13	11	10	0	0	0	0
city.ac.uk	66336	17	3056	1490	726	95	3660	1126	566	90
cranfield.ac.uk	9691	6	1555	1312	444	96	337	256	184	71
coventry.ac.uk	9575	15	543	435	332	87	1149	874	513	106
derby.ac.uk	11841	10	594	527	318	91	532	343	237	81
dmu.ac.uk	59917	33	2098	1386	750	98	2049	1688	859	94
dundee.ac.uk	51331	19	1981	1597	804	99	1930	1576	779	94
dur.ac.uk	77006	16	4075	2908	1053	97	5643	3423	1044	105
ed.ac.uk	229192	155	15699	10191	4184	104	10418	7944	3780	103

essex.ac.uk	52333	24	3029	2406	1245	98	2397	2027	992	99
ex.ac.uk	73500	17	3542	2957	943	102	4136	3431	1385	109
gcal.ac.uk	5551	26	756	501	349	80	1193	910	599	93
gla.ac.uk	367140	102	8342	5898	2557	103	11018	6757	2864	108
glam.ac.uk	14626	6	456	310	212	77	817	680	371	85
glos.ac.uk	1488	1	285	240	130	72	256	187	84	53
goldsmiths.ac.uk	9989	9	667	540	313	88	534	442	276	87
gre.ac.uk	21676	8	465	412	245	86	1879	1069	383	107
harper-adams.ac.uk	520	1	1384	363	33	28	1335	323	5	5
herts.ac.uk	63321	15	1116	878	490	97	3185	2060	661	101
hud.ac.uk	10918	9	937	558	269	87	885	768	333	91
hull.ac.uk	15619	12	2375	1386	558	100	1218	1013	446	94
hw.ac.uk	60396	25	4975	3779	1489	103	8696	5829	1946	108
ic.ac.uk	177225	96	21457	6479	3161	102	5817	3889	1677	95
kcl.ac.uk	59037	25	3388	2024	929	100	4121	3125	1097	98
keele.ac.uk	16809	6	1525	1187	524	99	1378	1200	546	100
king.ac.uk	13594	6	600	462	257	87	531	449	279	82
lamp.ac.uk	3497	4	313	258	137	65	332	248	120	56
lancs.ac.uk	71544	32	3386	2711	1234	102	4943	3609	1500	106
lboro.ac.uk	34203	23	1835	1527	701	100	2647	2152	785	101
le.ac.uk	34492	16	4054	2767	1052	98	3761	2981	1060	103
leeds.ac.uk	147827	70	10670	8725	2515	103	8781	6479	2720	107
lgu.ac.uk	13478	2	1732	1329	470	101	1986	1594	807	195
liv.ac.uk	18262	13	3706	2802	1042	100	1447	1146	491	84
livjm.ac.uk	30525	10	858	620	381	91	1165	907	448	99
lmu.ac.uk	10490	2	506	377	193	89	533	218	115	65
lse.ac.uk	29757	17	1763	1315	649	95	937	752	392	88
luton.ac.uk	1338	1	235	182	131	60	10	8	4	4
man.ac.uk	124093	94	8973	6425	2962	103	6109	4355	2091	100
mdx.ac.uk	41933	16	745	603	367	91	1480	1238	698	102

mmu.ac.uk	21108	23	1609	1136	626	98	2652	1926	1193	100
napier.ac.uk	98712	22	1623	1083	498	86	3737	2238	994	89
ncl.ac.uk	116941	50	4863	3759	1666	103	5318	4374	1571	96
newport.ac.uk	2949	17	128	91	67	43	243	209	144	71
northampton.ac.uk	6666	3	91	80	55	47	610	523	247	88
northumbria.ac.uk	13133	1	66674	2982	541	101	66460	2762	411	87
nott.ac.uk	70042	41	6784	3708	1674	101	4732	3332	1350	100
ntu.ac.uk	19491	24	1314	1014	643	101	1562	1158	700	103
open.ac.uk	45584	51	3636	2747	1660	102	2203	1793	1005	94
ox.ac.uk	197204	220	16021	11403	5719	104	16504	8140	4018	105
paisley.ac.uk	122	1	475	351	222	67	6	6	5	5
plym.ac.uk	23703	25	1031	866	595	96	1758	1447	904	99
port.ac.uk	70267	28	890	751	530	88	2507	1746	911	96
qmced.ac.uk	8804	10	487	257	135	56	599	466	300	74
qmw.ac.uk	23340	26	2860	2199	1138	92	1321	1073	633	80
qub.ac.uk	54601	28	2224	1520	848	94	3393	2522	1104	106
rdg.ac.uk	91278	33	4299	2978	1081	101	3484	2189	917	102
rgu.ac.uk	7589	3	756	594	300	93	1070	873	432	86
rhul.ac.uk	42738	25	1413	1051	655	96	2430	1788	965	95
salford.ac.uk	8246	38	1310	795	448	88	413	347	255	69
sbu.ac.uk	19067	16	1382	1002	536	101	4865	3188	848	93
shef.ac.uk	44142	19	6420	5124	1326	104	4589	2536	1048	98
shu.ac.uk	56375	14	1358	1055	445	101	1888	1328	476	100
soas.ac.uk	5681	3	430	361	180	74	448	356	161	52
soton.ac.uk	121726	105	18339	7439	2552	104	9749	6994	2467	108
st-and.ac.uk	83766	94	669805	72818	1754	94	668329	72372	1649	92
staffs.ac.uk	35756	8	842	682	356	90	1332	1034	498	101
stir.ac.uk	24312	39	1526	1253	645	97	1847	1423	809	108
strath.ac.uk	80482	73	3752	2665	1311	101	4018	2865	1502	96
sunderland.ac.uk	31277	11	1045	741	447	88	2326	1580	713	97

surrart.ac.uk	343	1	62	51	38	31	0	0	0	0
surrey.ac.uk	89091	23	3062	2418	1169	97	2750	2170	1040	98
susx.ac.uk	68030	27	3047	2406	1007	100	4303	2876	1056	96
swan.ac.uk	27521	25	1585	1205	544	99	2809	2072	735	100
tay.ac.uk	1131	6	246	205	143	57	60	33	31	23
tees.ac.uk	11174	12	644	393	238	79	494	372	271	97
tvu.ac.uk	7214	7	135	121	98	56	132	101	79	53
uce.ac.uk	21318	14	290	249	167	74	364	277	192	69
ucl.ac.uk	142018	90	13864	7961	2597	103	6775	5180	2420	100
uclan.ac.uk	12561	4	458	372	215	88	865	708	333	92
uea.ac.uk	46966	23	2374	1961	928	103	2139	1675	756	99
uel.ac.uk	17657	5	464	385	232	80	570	479	295	79
ukc.ac.uk	47112	11	2534	1992	932	99	2851	2099	854	106
ulh.ac.uk	1405	1	235	198	101	54	44	35	30	22
ulst.ac.uk	75727	32	1652	1254	703	100	2106	1431	696	102
umist.ac.uk	32922	50	2363	1673	984	98	2502	1552	751	82
unl.ac.uk	49112	9	0	0	0	0	0	0	0	0
uwe.ac.uk	24725	10	957	727	402	91	1505	1126	570	101
uwic.ac.uk	3971	6	192	125	80	47	211	174	116	51
warwick.ac.uk	47159	27	5179	4142	1632	104	3430	2714	1225	103
wlv.ac.uk	53346	11	2339	1879	946	103	4079	2416	1016	110
wmin.ac.uk	21174	13	764	620	371	95	636	539	337	78
worc.ac.uk	7184	3	131	112	80	48	724	607	198	85
york.ac.uk	67257	23	5460	3886	1394	102	5445	3891	1456	105

Jun 2003

Domain Name	Static Pages	Domains	Page inlinks	Directory inlinks	Domain inlinks	Site inlinks	Page outlinks	Directory outlinks	Domain outlinks	Site outlinks
abdn.ac.uk	123566	16	2447	1964	855	103	5116	3947	1007	111
aber.ac.uk	92137	16	1839	1406	649	101	3532	2537	1082	110
abertay.ac.uk	481	2	235	193	138	56	15	9	8	8
anglia.ac.uk	22915	15	1216	703	396	87	1126	926	561	100
aston.ac.uk	11734	16	993	709	436	87	1203	997	582	107
bangor.ac.uk	22265	19	1383	1013	578	96	1035	845	495	105
bath.ac.uk	73754	14	2957	2145	782	102	3630	2804	1126	106
bathspa.ac.uk	3071	5	155	139	95	58	522	466	294	107
bbk.ac.uk	29879	25	2011	1291	701	102	3421	2369	872	99
bham.ac.uk	252166	144	7147	5552	2143	107	9295	5723	2794	121
bournemouth.ac.uk	29092	16	694	527	309	93	613	498	348	88
brad.ac.uk	34704	18	1270	999	494	95	2164	1746	846	103
bris.ac.uk	121651	68	6170	4703	2286	111	8377	6023	2678	114
brookes.ac.uk	25883	10	1301	1054	426	101	1302	969	411	96
brunel.ac.uk	41372	4	1602	1255	636	100	1469	1281	585	109
bton.ac.uk	43399	21	1263	977	545	101	14635	1258	579	98
buckingham.ac.uk	1459	1	51	48	36	31	19	19	19	14
cam.ac.uk	324795	270	45477	11979	5644	113	25217	14207	5794	118
cant.ac.uk	8393	25	250	211	145	71	511	437	276	79
cf.ac.uk	47246	20	3280	2543	1189	104	2959	2538	932	111
chichester.ac.uk	355	1	15	13	11	9	0	0	0	0
city.ac.uk	46493	17	3011	1470	709	100	1953	987	534	96
cranfield.ac.uk	13324	8	1283	1021	435	98	380	265	188	70
coventry.ac.uk	11205	13	577	472	337	90	1070	857	555	115
derby.ac.uk	10805	8	509	451	310	93	542	347	235	82
dmu.ac.uk	1271	3	1974	1296	733	102	33	19	11	7
dundee.ac.uk	68871	18	1950	1644	822	105	30508	1806	875	100

dur.ac.uk	94775	18	4234	2990	1054	103	4844	3129	1071	110
eca.ac.uk	226	1	110	93	63	37	0	0	0	0
ed.ac.uk	266073	166	15129	9827	4214	113	9681	7668	3998	119
essex.ac.uk	55126	19	2969	2253	1192	103	3012	2341	1047	107
ex.ac.uk	77766	19	3005	2465	942	107	4298	3460	1391	121
gcal.ac.uk	8232	31	775	555	395	87	1077	869	597	100
gla.ac.uk	136197	101	8417	5867	2633	109	11196	7456	3365	119
glam.ac.uk	10379	3	492	334	226	82	743	605	328	88
glos.ac.uk	1509	1	310	273	140	79	264	194	88	56
goldsmiths.ac.uk	12697	11	1434	569	293	86	572	453	305	90
gre.ac.uk	23994	16	458	403	237	83	2705	1311	489	114
gsa.ac.uk	2406	1	168	139	67	37	84	67	32	22
harper-adams.ac.uk	911	1	8545	689	34	28	8498	648	4	4
health-homerton.ac.uk	227	1	1319	58	10	4	1286	48	10	7
herts.ac.uk	21599	15	1144	884	480	96	1934	1529	691	108
hud.ac.uk	28834	6	720	530	281	90	937	781	322	93
hull.ac.uk	16909	17	1889	1407	559	103	1196	973	478	99
hw.ac.uk	55307	23	4888	3659	1442	110	7512	5037	1921	112
ic.ac.uk	312711	86	21423	6359	3134	109	5833	3913	1655	104
icr.ac.uk	1190	1	215	184	115	53	54	47	34	24
ioe.ac.uk	2472	4	657	492	276	82	177	135	95	52
kcl.ac.uk	44471	26	3663	2174	983	109	3790	2797	1047	107
keele.ac.uk	18601	6	1456	1169	529	101	2261	1426	614	109
king.ac.uk	6384	5	590	451	273	91	273	224	156	70
lamp.ac.uk	3997	3	306	256	135	56	505	413	198	101
lancs.ac.uk	79894	38	3408	2678	1230	109	4924	3706	1635	117
lboro.ac.uk	48274	21	2149	1730	763	108	1662	1300	695	108
le.ac.uk	40386	17	4160	2830	1052	104	3646	2895	1076	110
leeds.ac.uk	118518	78	11569	9321	2574	115	8593	6318	2698	117

lgu.ac.uk	11956	1	1677	1240	441	104	1693	1361	672	195
lincoln.ac.uk	13952	3	218	174	114	65	82	76	61	39
liv.ac.uk	20355	5	4081	2947	1088	106	929	735	240	70
livjm.ac.uk	37614	12	867	558	361	91	1277	1014	502	107
lmu.ac.uk	8213	2	460	387	197	88	285	204	112	64
london.edu	4657	4	87	80	54	32	40	38	30	22
lse.ac.uk	36929	20	1913	1487	743	101	1145	957	491	99
lshtm.ac.uk	3920	7	442	294	155	58	357	316	184	64
luton.ac.uk	1422	2	236	176	124	64	19	19	19	16
man.ac.uk	130099	127	10262	6702	3029	107	7824	5613	2719	114
mdx.ac.uk	36057	15	678	546	344	94	1147	910	515	101
mmu.ac.uk	27604	25	1656	1171	637	102	2174	1380	750	96
napier.ac.uk	46387	21	1428	965	485	90	2959	1677	862	99
ncl.ac.uk	112439	54	4851	3688	1688	110	4674	3784	1800	108
newport.ac.uk	3731	15	105	79	61	38	362	295	195	77
northampton.ac.uk	8435	5	116	101	71	54	656	532	256	92
northumbria.ac.uk	38007	1	145343	4532	570	106	145292	4194	428	92
nott.ac.uk	74639	44	9260	3850	1678	106	5503	3618	1393	109
ntu.ac.uk	19719	29	1313	1022	692	106	1249	982	692	114
open.ac.uk	43439	46	3680	2798	1704	107	4487	2264	1105	109
ox.ac.uk	227224	236	16608	11676	5894	112	17229	8560	4134	120
paisley.ac.uk	136	1	468	336	208	63	6	6	5	5
plym.ac.uk	16338	23	1102	898	618	100	1616	1318	793	104
port.ac.uk	85060	25	902	784	537	91	2835	1987	858	106
qmced.ac.uk	10746	10	347	245	149	59	691	549	349	83
qmw.ac.uk	22209	23	2597	1971	1026	95	865	669	432	88
qub.ac.uk	51818	34	1953	1587	835	93	3367	2539	1199	116
rca.ac.uk	3940	3	146	128	89	51	39	35	21	18
rdg.ac.uk	94207	37	4017	2832	1063	103	3501	2139	1080	111
linst.ac.uk	1324	1	2968	371	196	76	2789	213	78	35

rgu.ac.uk	5567	3	941	595	305	94	779	662	333	86
rhul.ac.uk	23801	23	1645	1129	683	98	2328	1700	878	101
roehampton.ac.uk	5267	1	254	197	111	70	331	270	132	72
rvc.ac.uk	768	1	306	90	41	27	26	21	14	12
salford.ac.uk	18232	48	1674	850	507	90	809	655	485	95
sbu.ac.uk	20896	11	1310	949	525	104	4756	3260	817	99
sghms.ac.uk	4881	6	294	240	157	66	232	204	121	51
shef.ac.uk	101744	25	6480	5067	1354	110	7930	4913	1453	120
shu.ac.uk	53085	15	1527	1177	470	102	2031	1379	561	105
soas.ac.uk	5802	3	518	422	204	73	385	276	141	50
soton.ac.uk	199815	114	19683	7879	2664	110	7670	5551	2437	120
st-andrews.ac.uk	36157	30	329282	56200	1504	101	325397	54176	716	87
staffs.ac.uk	42424	8	821	674	355	95	1220	938	466	106
stir.ac.uk	24503	50	1500	1232	648	99	1489	1186	759	121
strath.ac.uk	59485	86	3323	2395	1238	106	3546	2507	1411	99
sunderland.ac.uk	22308	9	924	711	432	93	2266	1540	700	104
surrart.ac.uk	626	1	50	45	39	31	26	16	7	7
surrey.ac.uk	96455	23	3065	2388	1167	101	2840	2194	1062	104
susx.ac.uk	68556	21	3225	2536	1043	109	4070	2819	1044	105
swan.ac.uk	23190	24	1883	1351	560	98	2541	1993	731	104
tees.ac.uk	14537	12	505	372	243	82	460	329	249	94
tvu.ac.uk	11318	14	149	133	112	55	197	158	123	62
uce.ac.uk	12693	20	366	272	184	79	534	430	287	93
ucl.ac.uk	199796	114	13421	8701	2784	112	9049	6755	3101	112
uclan.ac.uk	13793	4	475	383	216	88	761	558	270	89
uea.ac.uk	60261	25	2496	1961	908	104	1782	1537	776	105
uel.ac.uk	20951	3	422	357	224	82	639	518	296	89
ukc.ac.uk	32351	9	2583	1960	901	106	2643	2053	809	116
ulsop.ac.uk	104	1	36	35	25	19	3	3	2	2
ulst.ac.uk	95960	26	1760	1196	687	102	2286	1594	776	110

umist.ac.uk	55477	52	2688	1859	1070	102	4035	2001	829	91
unl.ac.uk	25023	8	0	0	0	0	0	0	0	0
uwe.ac.uk	26183	12	1091	801	440	96	1215	874	462	98
uwic.ac.uk	4336	3	207	155	91	54	182	148	87	59
warwick.ac.uk	61017	24	5452	4196	1655	111	3649	2933	1347	114
wlv.ac.uk	34974	13	2242	1760	936	112	3923	2300	1013	122
wmin.ac.uk	30699	12	690	572	340	97	617	532	327	88
worc.ac.uk	9088	2	240	141	96	54	728	606	203	89
york.ac.uk	88844	23	5590	3922	1427	108	4912	3769	1452	112

Jun 2004

Domain Name	Static Pages	Domains	Page inlinks	Directory inlinks	Domain inlinks	Site inlinks	Page outlinks	Directory outlinks	Domain outlinks	Site outlinks
abdn.ac.uk	133841	14	2339	1831	795	102	4958	3869	1044	112
aber.ac.uk	111930	14	1698	1263	585	106	3343	2379	979	113
anglia.ac.uk	20400	9	1058	600	337	86	812	634	317	93
arts.ac.uk	3796	2	284	211	155	71	89	79	65	49
aston.ac.uk	13110	15	860	624	381	83	1114	935	514	103
bangor.ac.uk	24095	22	1471	964	562	94	1277	912	572	107
bath.ac.uk	64446	10	2572	1940	719	100	3436	2701	1134	105
bathspa.ac.uk	3972	4	136	120	78	50	474	370	254	100
bbk.ac.uk	40292	28	1875	1254	684	99	4463	2464	919	100
bham.ac.uk	278841	160	6394	4990	2146	111	9138	4579	2423	123
bournemouth.ac.uk	28439	19	698	501	275	90	580	480	337	89
brad.ac.uk	34168	17	1090	866	445	99	2156	1731	825	102
bris.ac.uk	168002	68	5761	4450	2121	111	7645	5561	2616	116
brookes.ac.uk	28785	14	1340	1051	424	103	1145	910	433	99
brunel.ac.uk	36858	2	1367	1100	564	101	1075	930	384	110
bton.ac.uk	18332	18	1138	879	479	99	885	713	416	102
buckingham.ac.uk	1479	1	45	39	30	27	21	21	21	17

cam.ac.uk	330336	270	44685	11352	5410	113	21223	11443	5351	118
cant.ac.uk	7067	27	272	210	145	67	556	472	299	84
cf.ac.uk	47269	24	2725	2120	1006	107	2999	2456	883	110
chichester.ac.uk	489	0	15	13	10	8	0	0	0	0
city.ac.uk	45647	16	1957	1309	664	99	955	800	495	99
coventry.ac.uk	11448	12	1423	445	321	84	751	637	412	115
cranfield.ac.uk	15991	9	1110	828	402	95	385	263	179	68
derby.ac.uk	7677	6	429	361	245	82	252	216	145	62
dmu.ac.uk	1042	3	1909	1175	654	97	7	7	6	5
dundee.ac.uk	74498	19	1906	1536	784	104	30379	1723	868	102
dur.ac.uk	72673	17	4094	2941	1034	103	4397	2751	996	98
eca.ac.uk	76	1	85	75	53	30	8	8	6	6
ed.ac.uk	224629	178	13151	8686	3945	111	9360	7436	3773	115
essex.ac.uk	49210	18	2833	2103	1099	101	2721	2167	928	110
ex.ac.uk	88130	20	2872	2238	884	107	4210	3303	1386	122
gcal.ac.uk	6887	24	806	497	357	84	667	529	388	95
gla.ac.uk	149117	88	8086	5533	2512	111	9137	6349	2817	115
glam.ac.uk	8545	3	484	316	205	74	665	531	298	92
glos.ac.uk	5673	2	327	281	141	80	338	223	105	67
goldsmiths.ac.uk	8925	11	578	464	296	88	355	308	237	79
gre.ac.uk	17769	13	408	372	225	84	1833	1054	456	115
gsa.ac.uk	27	0	148	122	66	34	0	0	0	0
harper-adams.ac.uk	831	1	52	41	30	24	4	4	4	4
health-homerton.ac.uk	522	2	57	31	9	3	108	77	41	7
herts.ac.uk	20836	18	1073	798	449	93	1630	1372	644	109
hope.ac.uk	3528	5	102	79	59	41	272	264	156	77
hud.ac.uk	10898	4	907	582	282	90	650	518	236	82
hull.ac.uk	18304	12	1707	1196	510	101	1077	895	445	101
hw.ac.uk	55126	24	4326	3258	1270	108	6448	4562	1844	111

ic.ac.uk	107762	87	7634	5696	2893	110	5071	3566	1662	103
icr.ac.uk	1457	1	208	170	101	49	47	39	29	22
ioe.ac.uk	2413	4	641	484	279	84	240	169	116	62
kcl.ac.uk	52884	21	4469	2160	945	107	10930	2739	956	103
keele.ac.uk	20559	6	1271	1068	494	101	1643	1396	589	109
kent.ac.uk	39822	10	2554	1782	815	105	3348	2238	934	118
king.ac.uk	6275	5	555	419	250	89	374	332	193	90
lamp.ac.uk	2961	2	291	235	116	53	300	178	105	51
lancs.ac.uk	80251	38	3268	2550	1171	110	4442	3292	1470	122
lboro.ac.uk	38682	29	2326	1707	728	107	1707	1321	708	107
le.ac.uk	49352	21	3924	2637	953	109	4169	3064	1177	111
leeds.ac.uk	107340	86	10701	8408	2492	117	8695	6368	2628	118
lincoln.ac.uk	13491	5	191	155	98	60	305	212	121	66
liv.ac.uk	29672	13	3723	2675	1035	105	1300	1030	398	84
livjm.ac.uk	14545	11	840	545	360	90	1044	831	407	112
lmu.ac.uk	10507	2	488	375	193	84	302	154	79	46
london.edu	2435	3	122	95	62	36	38	36	31	24
londonmet.ac.uk	40276	12	1428	1043	415	105	1116	894	492	97
lsbu.ac.uk	26893	15	1186	846	479	102	4836	3337	920	94
lse.ac.uk	44067	25	2001	1522	723	101	1328	1047	565	99
lshtm.ac.uk	4053	6	416	301	170	56	361	324	166	57
luton.ac.uk	1742	2	175	154	101	59	10	10	10	9
man.ac.uk	107806	144	11272	6819	2861	109	8427	5492	2737	116
mdx.ac.uk	31710	16	720	563	333	89	1381	1152	577	104
mmu.ac.uk	25977	37	1515	1042	593	100	1661	931	558	90
napier.ac.uk	33921	21	1243	921	450	84	2263	1189	724	98
ncl.ac.uk	140036	59	4448	3374	1578	111	5917	4203	2137	111
newport.ac.uk	4202	16	304	101	57	33	257	174	126	51
northampton.ac.uk	6575	3	96	80	61	48	247	177	104	66
northumbria.ac.uk	22033	8	1195	887	451	103	1076	765	416	91

nott.ac.uk	87679	31	10407	3911	1592	105	6419	3681	1284	109
ntu.ac.uk	18909	29	1231	948	640	107	1340	1096	739	115
open.ac.uk	48167	53	3378	2579	1559	110	4214	2249	1257	107
ox.ac.uk	275784	243	21449	10663	5548	113	11854	8256	4115	117
paisley.ac.uk	2611	8	434	304	188	60	326	260	160	54
plym.ac.uk	11440	13	1037	832	554	97	1427	1176	624	101
port.ac.uk	91555	25	829	726	497	90	2815	1652	739	98
qmced.ac.uk	19165	8	358	248	155	60	591	520	343	82
qmw.ac.uk	14580	21	2641	2096	1092	94	1179	914	430	78
qub.ac.uk	54120	40	1791	1465	814	98	2279	1776	926	117
rca.ac.uk	6860	2	156	125	82	46	40	37	23	20
rdg.ac.uk	74695	44	3726	2680	1049	106	2298	1676	895	105
rgu.ac.uk	5276	4	755	514	261	89	685	572	293	83
rhul.ac.uk	25725	23	1639	1099	659	95	1626	1085	643	97
roehampton.ac.uk	9272	1	252	168	86	56	264	190	104	63
rvc.ac.uk	945	1	309	90	38	23	25	22	13	10
salford.ac.uk	13998	39	1277	764	490	86	683	532	386	87
sghms.ac.uk	2787	5	307	236	143	61	166	110	69	35
shef.ac.uk	108667	29	6008	4764	1302	114	8477	5059	1466	122
shu.ac.uk	31486	10	1513	1030	457	101	1651	1083	566	105
soas.ac.uk	4287	3	564	405	197	68	282	170	83	42
soton.ac.uk	136941	123	15055	7367	2638	113	5943	4219	2103	121
staffs.ac.uk	83017	10	791	639	318	89	835	687	366	96
st-andrews.ac.uk	39828	40	4650	2771	1245	97	2372	1845	805	88
stir.ac.uk	27101	50	1473	1195	656	97	1650	1288	729	91
strath.ac.uk	54644	80	3237	2319	1194	106	3070	2154	1240	96
sunderland.ac.uk	25509	11	823	620	378	92	2354	1420	615	103
surrart.ac.uk	1056	2	49	43	36	27	31	21	12	12
surrey.ac.uk	97550	24	2842	2210	1106	104	2533	1894	977	101
susx.ac.uk	61197	23	3110	2439	991	110	3038	1974	918	99

swan.ac.uk	30525	25	1607	1223	559	100	2570	2081	793	103
tay.ac.uk	1565	6	225	187	134	51	48	41	40	30
tees.ac.uk	9592	12	589	372	247	79	288	238	192	95
tvu.ac.uk	14026	12	127	111	93	49	212	188	141	67
uce.ac.uk	25541	19	345	268	177	74	587	396	262	88
ucl.ac.uk	230330	111	11616	6661	2638	113	9082	6596	3029	113
uclan.ac.uk	16398	2	439	347	194	81	569	432	227	83
uea.ac.uk	57255	20	2243	1689	814	99	1487	1249	644	97
uel.ac.uk	23392	2	347	295	190	77	571	441	274	89
ulsop.ac.uk	1386	2	36	35	25	19	33	27	27	19
ulst.ac.uk	61870	26	1759	1084	616	104	2319	1555	733	110
umist.ac.uk	44252	44	3396	1859	998	100	5151	2337	723	86
uwe.ac.uk	29688	17	1045	762	435	89	1231	860	462	104
uwic.ac.uk	5336	3	191	141	87	51	194	153	95	64
warwick.ac.uk	77267	24	5154	3867	1591	109	4019	3234	1273	109
wlv.ac.uk	20039	10	2181	1566	851	109	3723	2092	844	122
wmin.ac.uk	22320	11	641	522	321	96	694	473	318	80
worc.ac.uk	29343	4	177	132	84	57	2549	552	249	90
york.ac.uk	103317	21	5365	3669	1344	108	4805	3640	1432	108

Jul 2005

Domain Name	Static Pages	Domains	Page inlinks	Directory inlinks	Domain inlinks	Site inlinks	Page outlinks	Directory outlinks	Domain outlinks	Site outlinks
abdn.ac.uk	100507	14	2235	1742	769	96	2611	2236	776	87
aber.ac.uk	109577	15	1417	1128	542	99	2954	2042	859	102
apu.ac.uk	36250	10	566	450	291	85	949	705	321	92
arts.ac.uk	5850	3	278	193	138	63	92	81	69	49
aston.ac.uk	10899	9	779	553	353	81	711	604	352	84
bangor.ac.uk	55562	23	1317	831	482	90	1248	789	442	101
bath.ac.uk	75864	0	2597	1887	700	92	3812	3004	1279	102

bathspa.ac.uk	5724	5	157	125	79	45	459	380	252	97
bbk.ac.uk	38578	34	1750	1167	635	86	4417	2323	881	87
bham.ac.uk	128184	177	5978	4712	2044	102	8813	4333	2346	110
bournemouth.ac.uk	35560	19	661	502	265	88	616	521	367	92
brad.ac.uk	33002	15	997	802	404	93	1026	874	422	93
bris.ac.uk	199552	64	5708	4325	1977	104	7507	5472	2303	105
brookes.ac.uk	27683	14	1231	991	413	95	1009	801	386	90
brunel.ac.uk	14878	4	1340	1046	492	94	600	506	309	76
bton.ac.uk	17747	17	1063	810	451	92	811	641	379	96
buckingham.ac.uk	1638	1	45	41	32	25	25	25	23	19
cam.ac.uk	350371	277	44146	11042	5216	101	19910	10692	5064	105
canterbury.ac.uk	11366	34	257	193	135	63	639	534	340	80
cardiff.ac.uk	32118	10	2674	2005	954	100	1761	1413	489	90
chichester.ac.uk	571	0	14	12	9	7	0	0	0	0
city.ac.uk	58533	16	2162	1151	607	87	990	687	432	81
coventry.ac.uk	8144	11	553	431	295	83	712	619	423	107
derby.ac.uk	5705	4	383	328	227	80	270	180	122	65
dmu.ac.uk	34664	17	1785	1061	557	92	745	535	304	78
dundee.ac.uk	82595	19	1763	1433	719	95	30281	1613	808	96
dur.ac.uk	76909	17	4054	2873	1017	92	3885	2529	886	91
ed.ac.uk	296994	260	11076	8117	3701	100	10547	8251	4386	105
essex.ac.uk	60369	20	2855	2050	1072	97	2864	2235	967	102
ex.ac.uk	82299	35	2657	2084	867	100	3678	2685	1397	109
gcal.ac.uk	16532	23	859	511	358	83	1131	892	627	94
gla.ac.uk	161440	98	7819	5270	2410	103	9029	5672	2498	101
glam.ac.uk	9470	9	452	284	178	68	698	457	203	75
glos.ac.uk	6819	2	299	258	142	74	364	287	156	75
goldsmiths.ac.uk	9340	13	606	496	286	77	427	346	256	71
gre.ac.uk	18375	12	474	362	203	77	1101	691	408	108
harper-adams.ac.uk	919	1	43	39	26	20	12	12	12	12

health-homerton.ac.uk	863	2	70	40	12	7	157	107	59	21
herts.ac.uk	15128	14	1071	777	456	89	1231	998	438	99
hope.ac.uk	4308	4	105	75	58	39	296	263	152	71
hud.ac.uk	14039	8	1085	597	273	88	567	472	254	83
hull.ac.uk	18824	12	1993	1624	478	94	1142	952	476	93
hw.ac.uk	69802	26	3988	2980	1170	102	6615	4676	1883	104
ic.ac.uk	115449	59	7231	5242	2647	99	4333	2787	1240	86
kcl.ac.uk	54008	18	4118	2159	900	97	11694	2818	941	95
keele.ac.uk	21307	6	1132	925	439	92	1791	1425	592	105
kent.ac.uk	47507	10	2560	1822	744	99	3638	2303	905	106
king.ac.uk	8214	4	477	350	213	78	317	286	173	75
lamp.ac.uk	2802	2	266	210	95	45	306	188	110	51
lancs.ac.uk	103395	43	3279	2518	1115	101	3742	2850	1362	100
lboro.ac.uk	43439	23	2187	1634	674	100	1809	1347	711	99
le.ac.uk	51574	19	3654	2535	911	102	4216	3005	1102	104
leeds.ac.uk	181382	88	10607	8171	2411	106	8369	6427	2504	108
lincoln.ac.uk	2803	5	166	140	88	53	137	91	50	32
liv.ac.uk	41321	6	3509	2511	992	100	1107	839	278	81
livjm.ac.uk	27018	11	686	500	331	86	1118	948	413	105
lmu.ac.uk	8788	1	411	327	171	80	184	103	50	32
londonmet.ac.uk	43675	8	1821	1495	382	100	864	765	391	92
lsbu.ac.uk	28436	15	1007	714	395	92	4878	3395	935	92
lse.ac.uk	51512	25	2082	1592	721	92	1445	1071	569	82
luton.ac.uk	4627	5	128	120	82	47	132	130	69	34
manchester.ac.uk	129421	133	8404	6239	3173	105	7727	5046	2602	104
mdx.ac.uk	23099	17	676	545	307	83	1422	1119	529	94
mmu.ac.uk	25992	42	1201	938	553	94	1480	821	480	86
napier.ac.uk	21974	18	1086	811	428	81	1206	815	509	93
ncl.ac.uk	98735	53	3985	3120	1432	103	4396	3405	1598	102

newport.ac.uk	6379	19	265	101	55	34	320	239	182	70
northampton.ac.uk	13178	6	99	84	59	40	589	472	227	85
northumbria.ac.uk	5308	2	1086	802	425	99	216	165	94	50
nott.ac.uk	86856	38	10411	3825	1488	99	5135	4512	1176	99
ntu.ac.uk	10737	10	1100	849	565	97	529	439	336	99
open.ac.uk	51403	53	3122	2411	1451	103	2538	1942	1175	96
ox.ac.uk	293597	255	21112	10102	5288	104	13078	8726	4124	106
paisley.ac.uk	4067	13	421	290	176	52	415	336	203	62
plymouth.ac.uk	5	0	1014	778	504	91	0	0	0	0
port.ac.uk	100514	23	723	630	442	83	2338	1550	698	94
qmced.ac.uk	11286	10	241	195	130	52	360	325	238	64
qmul.ac.uk	81706	41	2917	1966	999	90	2349	1810	849	88
qub.ac.uk	60898	42	1657	1352	757	89	2916	1968	910	107
rdg.ac.uk	60686	47	3456	2460	1017	98	2203	1612	862	95
rgu.ac.uk	4627	3	673	501	250	83	308	238	159	65
rhul.ac.uk	43364	28	1713	1062	602	87	2096	1517	880	94
salford.ac.uk	15508	42	800	653	437	81	658	504	362	85
shef.ac.uk	117345	33	5523	4439	1230	106	6455	4897	1530	102
shu.ac.uk	28581	9	1272	924	450	99	923	708	401	86
soas.ac.uk	2075	2	569	399	184	65	34	21	16	10
soton.ac.uk	161640	133	12193	6486	2489	105	5495	4116	2051	108
staffs.ac.uk	74628	11	783	618	295	86	712	572	322	91
st-andrews.ac.uk	43563	46	4677	2691	1236	86	2520	2011	936	84
stir.ac.uk	24383	59	1465	1133	624	92	1541	1176	694	84
strath.ac.uk	58530	77	3386	2263	1205	98	3142	2031	1075	93
sunderland.ac.uk	22643	11	682	506	330	86	2091	1207	521	98
surrart.ac.uk	2937	2	44	37	29	22	30	25	23	22
surrey.ac.uk	59755	28	2523	1991	1027	99	2139	1649	855	93
sussex.ac.uk	85719	22	2945	2259	904	101	3021	1962	890	91
swan.ac.uk	44980	28	1483	1126	532	89	2918	2261	851	97

tay.ac.uk	2781	10	216	180	137	44	153	80	76	48
tees.ac.uk	2190	5	428	302	202	72	39	36	27	20
tvu.ac.uk	3365	6	124	104	84	43	94	90	76	48
uce.ac.uk	17184	14	341	264	177	68	421	243	173	72
ucl.ac.uk	231286	105	11082	6665	2587	104	9479	6787	2943	101
uclan.ac.uk	21084	3	428	335	197	79	694	515	257	83
uea.ac.uk	55504	20	2172	1617	767	97	1449	1195	654	91
uel.ac.uk	24387	3	318	267	171	74	614	451	266	84
ulst.ac.uk	90638	26	1507	913	538	97	2567	1489	697	103
uwe.ac.uk	29355	23	1085	752	429	89	1416	950	495	99
uwic.ac.uk	5591	3	191	135	81	47	205	150	96	63
warwick.ac.uk	92588	22	4884	3848	1497	98	4174	3380	1228	103
wlv.ac.uk	32105	12	1636	1288	757	103	3865	2298	836	110
wmin.ac.uk	19881	6	585	475	291	88	475	359	247	79
worc.ac.uk	4242	4	143	116	80	48	405	346	161	73
york.ac.uk	107867	21	5392	3780	1262	100	4780	3511	1343	99

Appendix 6: Randomly Selected Link Data for New Zealand Universities 2000

Page	Link	Pilot Study	Primary Motivation	Cross-check
www.auckland.ac.nz/lbr/geol/geolorg.htm	www.csn.net/~tbrez/cogs/index.html		Professional	
www.ctru.auckland.ac.nz/studies/progress/intro.html	www.hrc.govt.nz		Administrative	
www.auckland.ac.nz/lbr/anthro/105203.htm	anthro.AnnualReviews.org/current.shtml		Research Oriented	
www.auckland.ac.nz/lbr/anthro/antgatenzp.htm	online.anu.edu.au/caepr/		Research Oriented	Research Oriented
www.arts.auckland.ac.nz/eur/spanish/splitLA.html	cvax.ipfw.indiana.edu/~jehle/poetry.html		Professional	
www.auckland.ac.nz/lbr/geog/geogmeta.htm	faculty.washington.edu/krumme/resources/resources_abc.html		Technical	
www.auckland.ac.nz/lbr/maori/maosoft.htm	www.geocities.com/HotSprings/Bath/3034/takhome.htm		Technical	
www.law.auckland.ac.nz/itlaw/ITInfo.htm	www.austlii.edu.au/au/other/unswlj/thematic/1998 /vol21no2/greenleaf.html		Research Oriented	
www.ele.auckland.ac.nz/students/orange/interest.html	www.infotech.co.nz		Technical	
www.cs.auckland.ac.nz/~pgut001/links.html	www.cycon.com/		Technical	
www.law.auckland.ac.nz/itlaw/ITInfo.htm	entertainment.msn.com/news/eonline/0601/jroberts.asp		Personal	
www.auckland.ac.nz/lbr/libcats.htm	www.uq.edu.au/~mljeast/		Technical	Technical
www.auckland.ac.nz/lbr/eng/resources/netpsorg.htm	www.cpo.cn.net/		Administrative	
www.planning.auckland.ac.nz/info/PlanWrld.html	www.lib.berkeley.edu/ENVI/cityweb.html	Professional	Professional	
www.aut.ac.nz/news/navigator/	http://www.aut.ac.nz/corp/news/navigator/		Navigational	
www.aut.ac.nz/services/	.ait.ac.nz/corp/news/index.shtml		Personal	
www.aut.ac.nz/depts/researchoffice/textic/proj2.shtml	.ait.ac.nz/corp/aboutait/contact.shtml		Personal	
www.aut.ac.nz/stats/1997/usage_199710.html	.ait.ac.nz/depts/shrs/		Professional	Other
www.aut.ac.nz/depts/commstud/theory/wk8.htm	rheingold.com/vc/book/		Research Oriented	
www.library.canterbury.ac.nz/com/econ/econ_web.shtml	econwpa.wustl.edu/EconFAQ/EconFAQ.html		Professional	
www.pols.canterbury.ac.nz/ECSANZ/online.htm	.ecdel.org.au		Professional	
www.soci.canterbury.ac.nz/linksf.htm	.ibd.nrc.ca/~mansfield/feminism/		Research Oriented	Research Oriented
www.cosc.canterbury.ac.nz/~bim20/	.uic.edu/depts/psch/ohlson-1.html		Personal	

www.blds.canterbury.ac.nz/pcmag/roboty.html	.pcmag.com/discuss.htm/		Personal	
www.cosc.canterbury.ac.nz/~mpj17/204/unix4.html	nix.tmk.auckland.ac.nz/LDP/HOWTO /Emacs-Beginner-HOWTO.html	Technical	Technical	
www.cosc.canterbury.ac.nz/help/texmf/doc/html /catalogue/ctt.html	theory.uwinnipeg.ca/scripts/CTAN/support/tex2ltx.tar.gz		Technical	
www.unplugged.canterbury.ac.nz/training/detail.htm	.cs.waikato.ac.nz/cs/Staff/ian-h.-witten.html		Personal	
www.cosc.canterbury.ac.nz/help/texmf/doc/html /manpages/tth.html	venus.pfc.mit.edu/tth/Xfonts.html		Technical	
www.lincoln.ac.nz/libr/dbases/access.htm	agecon.lib.umn.edu		Educational	
www.lincoln.ac.nz/libr/law.htm	.law.indiana.edu/glsj/glsj.html		Research Oriented	Research Oriented
www.lincoln.ac.nz/comm/subjects/bmkt326/coke.htm	.nytimes.com/yr/mo/day/		Professional	
www.lincoln.ac.nz/systanz/page4.htm	.vuw.ac.nz/index.shtml	Administrative	Administrative	
www.massey.ac.nz/staff/	.aus.ac.nz		Professional	
www.massey.ac.nz/~wwits/services/www/wwwdev /html_cgi_ref.html	.rpi.edu/~decemj/pages/table.html		Educational	
www.IFS.massey.ac.nz/maple.htm	.math.scarolina.edu/~meade/math242/		Educational	
www-ist.massey.ac.nz/csnotes/201/lectures/index.html	.oreilly.com/		Educational	
www-ist.massey.ac.nz/csnotes/355/extras/links.html	.gnu.org/philosophy/philosophy.html		Technical	
www.IFS.massey.ac.nz/mathnews/NZMS68/positions.html	.math.auckland.ac.nz/~conder/NZMS/	Professional	Professional	
www.fims-www.massey.ac.nz/~is/papers/157796d.html	.adobe.com/products/acrobat/readstep.html		Technical	
www.massey.ac.nz/~cprichar/hecu2.html	.lancs.ac.uk/		Administrative	Administrative
www.massey.ac.nz/~i75202/projects/sw/outline.htm	.informatik.umu.se/~rwhit/ObjRedCog.html	Educational	Educational	
www-ist.massey.ac.nz/~audiogph/htdocs/links.html	.ee.surrey.ac.uk/Research/DLearn/		Technical	
www.library.otago.ac.nz/	.chmeds.ac.nz/services/library		Technical	
www.library.otago.ac.nz/guides/nz-resources.html	.forestresearch.cri.nz/		Research Oriented	
www.cs.otago.ac.nz/nnweb/FAQ4.html	.cs.utoronto.ca/DCS/People/Faculty/hinton.html		Personal	
www.healthsci.otago.ac.nz/division/medicine/sportsmed /home.html	healthsci/division/home.html		Administrative	
www.physed.otago.ac.nz/apa/main.html	info.lut.ac.uk/research/paad/home.html		Professional	

www.otago.ac.nz/DeepSouth/vol3no1/editorial.html	point.lycos.com/categories/		Other	
www.library.otago.ac.nz/guides/nz-stats.html	.who/whosis/		Administrative	
www.otago.ac.nz/cure/people.html	.abdn.ac.uk/public_health/hsru/staff/granta.htm		Personal	
www.divcom.otago.ac.nz/sirc/webpages/Conferences /SIRC97/body_sirc97.htm	.leeds.ac.uk		Administrative	Administrative
www.nzdis.otago.ac.nz/servlets/Dis/about.html	.icair.iac.org.nz/		Research Oriented	
www.vuw.ac.nz/index.shtml	.vuw.ac.nz/home/staff/index.html		Personal	
www.vuw.ac.nz/library/ejournals/ejj.html	.vuw.ac.nz/library/ejournals/ejt.html	Research Oriented	Research Oriented	
www.vuw.ac.nz/classics/drug.html	.sori.org/gayo/		Personal	
www.vuw.ac.nz/dlis/courses/533/m3srchtl.htm	.lincoln.ac.nz/libr/nz/	Technical	Technical	
www.mcs.vuw.ac.nz/courses/COMP305/2000 /LectureNotes/1.Introduction/tsld008.htm	.mcs.vuw.ac.nz/courses/COMP305/2000 /LectureNotes/1.Introduction/tsld008.htm		Superficial	
www.waikato.ac.nz/library/	waikato.ac.nz/library/distance/mba.html		Educational	
www.waikato.ac.nz/wfass/subjects/anthropology /research.shtml	waikato.ac.nz/		Administrative	
www.mngt.waikato.ac.nz/research/publications/editors.asp	.multi.demon.co.uk/journals.htm		Research Oriented	
www.mngt.waikato.ac.nz/depts/sml/journal/editors.htm	.umist.ac.uk/		Administrative	
www.c14.sci.waikato.ac.nz/nzaa/nzwww.html	.ccc.govt.nz/Library/		Technical	
www.help.waikato.ac.nz/telephones/callpilot /record_name.shtml	http://help.waikato.ac.nz/telephones/callpilot/record_name.shtml		Technical	
www.list.waikato.ac.nz/archives/nznog/1999/07/thrd2.html	.oac.uci.edu/indiv/ehood/mhonarc.html		Superficial	
www.cs.waikato.ac.nz/~ihw/index.html	.cpsc.ucalgary.ca/		Administrative	
www.cs.waikato.ac.nz/~marku/424/	.comlab.ox.ac.uk/archive/z.html		Educational	
www.cs.waikato.ac.nz/~rhl/spim.html	binger.centre.edu/classes/css21-winter96/spim/spim.html		Educational	
www.aut.ac.nz/services/	.ait.ac.nz/corp/aboutait/library.shtml		Technical	
www.aut.ac.nz/depts/researchoffice/textic/proj1.shtml	.ait.ac.nz/corp/aboutait/search.shtml		Navigational	
www.canterbury.ac.nz/webfind/test.cfm	.minedu.govt.nz/tertiary/review/		Professional	Professional
www.afis.canterbury.ac.nz/Afis213/co21300.htm	.discoverJade.com/		Technical	

www.cape.canterbury.ac.nz/Archive/summary2.htm	.gensym.com/customerstories/lilly.html		Personal	
www.cape.canterbury.ac.nz/Archive/bungay/bungay.html	.eng.rpi.edu/dept/chem-eng/WWW/faculty/ChE.html		Professional	
www.lincoln.ac.nz/stusrvc/stuserv.htm	.winz.govt.nz/student/index.html	Educational	Educational	
www.massey.ac.nz/%7Ewwifs/physsource.htm	www-physics.mps.ohio-state.edu/		Professional	
www.is157321.massey.ac.nz/lectr_notes.html	officeupdate.microsoft.com/2000/downloadDetails /Ppview97.htm		Technical	
www.plant-protection.massey.ac.nz/resources /publications.htm	.hortnet.co.nz/publications/nzpps/proceeds.htm		Technical	
www.cblmm.massey.ac.nz/Miscellaneous.htm	.lancs.ac.uk/users/edres/research/csalt.html		Professional	
www.library.otago.ac.nz/services/CDROM.html	.umi.com/pqdauto		Technical	
www.divcom.otago.ac.nz/tourism/link.htm	.url.co.nz/nzl.html		Social/leisure	
www.otago.ac.nz/cure/people.html	.abdn.ac.uk/public_health/hsru/staff/brazzellim.htm		Personal	
www.vuw.ac.nz/home/faculties.html	magog.fca.vuw.ac.nz/	Administrative	Administrative	
www.scim.vuw.ac.nz/comms/courses/comm505 /schedule/..\links_strategy.htm	comminit.com/power_point/planning_strategy/sld001.htm		Research Oriented	
www.waikato.ac.nz/programmes.shtml	ecommerce.ac.nz		Educational	
www.mngt.waikato.ac.nz/depts/sml/	web.mit.edu/linguistics/www/chomsky.home.html		Personal	
www.installations.its.waikato.ac.nz/download.html	.adobe.com/prodindex/acrobat/readstep.html		Technical	Technical
www.mngt.waikato.ac.nz/depts/mnss/courses /ssm/index.htm	.orsoc.org.uk/home.html		Research Oriented	
www.list.waikato.ac.nz/archives/prir-l/1997/09 /msg00051.html	.uq.edu.au/gsm	Research Oriented	Research Oriented	
www.cs.waikato.ac.nz/~singlis/index.html	.amazon.com/exec/obidos/ASIN/1558605703/internz		General/Informative	
www.its.canterbury.ac.nz/nljul98/part3.htm	.compinfo.co.uk/y2k.htm		Technical	
www.massey.ac.nz/%7Ewwwlib/subjects/evolution.htm	.netscape.com		Technical	
www.massey.ac.nz/~KBirks/gender/factoids.htm	.calib.com/nccanch/pubs/stats.htm		Professional	
www.kel.otago.ac.nz/maaka/pukapuka.html	.te-kawerau-a-maki.iwi.nz/		Personal	
www.vuw.ac.nz/dlis/courses/533/m4srchtc.htm	.infotech.co.nz/		Technical	
www.waikato.ac.nz/nzcoll/nz14pacific.html	.nor.com.au/media/kmail/		Professional	

www.econ.canterbury.ac.nz/nzgov.htm	.consumer-ministry.govt.nz/		Administrative	
www.soci.canterbury.ac.nz/biograph/wirth.htm	.harlingen.tstc.edu/pages/soci/soci1301/c07text.htm		Technical	
www.massey.ac.nz/~KBirks/gender/menshealth.htm	menshealth.curtin.edu.au/		Personal	
www.physics.otago.ac.nz/research/bec2/vortex/	xxx.lanl.gov/abs/cond-mat/9902092		Research Oriented	Research Oriented
www.physics.otago.ac.nz/~justinb/	GuestWorld.Tripod.Lycos.com/		Personal	

Appendix 7: Randomly Selected Link Data for Australian Universities 2000

Page	Link	Pilot Study	Primary Motivation	Cross-check
www.adfa.oz.au/DOD/RAAF/	http://www.defence.gov.au/RAAF/		General/Informative	
www.adfa.oz.au/DOD/recruit/army/armcoff.htm	http://defencejobs.defence.gov.au/army/armcoff.htm		General/Informative	
anusf.anu.edu.au/hpc_visualization.html	www.nas.nasa.gov/TechnicalSummaries/technical_summaries.html		Technical	
msowww.anu.edu.au/~lisa/	www.travelaus.com.au/index.html		Personal	Social/leisure
www.ise.canberra.edu.au/computing	http://www.ise.canberra.edu.au/computing/		Administrative	Administrativ
beth.canberra.edu.au/Faculty/Tsrc.asp	http://www.ise.canberra.edu.au/SRC		Administrative	
beth.canberra.edu.au/Faculty/TITSrv.asp	http://www.ise.canberra.edu.au/CompServices	Navigational	Navigational	
www.library.cqu.edu.au/internet/communicate.htm	www.lib.berkeley.edu/TeachingLib/Guides/Internet/Glossary.html	Educational	Educational	
www.ahs.cqu.edu.au/~wangw/html/psyc/psycdept.html	www.washington.edu:1180/		Navigational	
library-gateway.cqu.edu.au/vtls/english/vtls-advanced.html	www.vtls.com		Superficial	Superficial
www.ahs.cqu.edu.au/~wangw/html/psyc/journal.html	www.apa.org/journals/cp.html		Professional	
www.csu.edu.au/links/act.html	203.37.72.50/parts/parts.html		Technical	
sin.csu.edu.au/sin/netpub/	www.ncsa.uiuc.edu/SDG/Docs/fill-out-forms/overview.html		Technical	
life.csu.edu.au/~dspennem/MCHMIS/NewCaledonia_Ovw.html	www.essi.fr/PubHTML/Nelle_Caledonie/english/nc_hist.html	Personal	Personal	
www.csu.edu.au/division/healsafe/library/links.htm	www.dir.ca.gov/		Administrative	
www.vc.curtin.edu.au/research/about/centres.html	ssda.anu.edu.au/ACSPRI/index.html		Professional	
lisweb.curtin.edu.au/staff/gwpersonal/searchtut/subject.html	omni.ac.uk/		Professional	
www.curtin.edu.au/curtin/dept/pharmacy/career/index.html	ii.net/~blueboy/design	Superficial	Superficial	
dali.ece.curtin.edu.au/notice/profsocs.html	www.acm.org		Research Oriented	
www.deakin.edu.au/stud/gas/links.htm	www.deetya.gov.au/jobguideonline/default.htm		Personal	
www.deakin.edu.au/career/career2_eps.html	www.macromedia.com		Technical	
www.deakin.edu.au/library/govinfo.html	www.aph.gov.au/hansard/		Professional	
www.deakin.edu.au/career/career2_eps1.html	www.nla.gov.au/oz/gov		Technical	
www.bio.flinders.edu.au/paa/pares.htm	wwwnt.scisoc.org/ismpmi/		Professional	
wwwehlt.flinders.edu.au/philosophy/MDavies/links.html	www.infoseek.com/		Personal	
www.ssn.flinders.edu.au/Politics/osites2.htm	info.dpac.tas.gov.au/features/ausconstitution.html		Professional	
www.jcu.edu.au/courses/info/musicedu.html	www.adobe.com/products/acrobat/readstep2.html		Technical	

www.es.jcu.edu.au/dept/Earth/schools/schools.shtml	www.abc.com.au/default.htm		Personal	
www.jcu.edu.au/courses/info/anthropo.html	www.adobe.com/products/acrobat/readstep2.html		Technical	
www.latrobe.edu.au/www/anzsms/Societies.html	www2.ifrn.bbsrc.ac.uk/bmss/		Administrative	
www.lib.latrobe.edu.au/AHR/goodo/home2.html	muse.jhu.edu	Research Oriented	Research Oriented	
www.ee.latrobe.edu.au/internal/links.html	www.wordperfect.com/		Technical	
www.latrobe.edu.au/www/chemistry/staff/ms.html	www.chemistry.unimelb.edu.au/ResHand2/Staff/WEDD.html		Personal	Research Oriented
www.lib.monash.edu.au/vl/thes/thessour.htm	www.lib.uwaterloo.ca/TUG/ETD/open/search.html		Technical	
www.monash.edu.au/campuses/gippsland/facilities.html	130.194.214.54/childcare.html		Educational	
www.monash.edu.au/commserv/info/selfhelp.html	aspen.uml.edu/www/counseling/suicide.html		Social/leisure	
www.student.mq.edu.au	www.muu.org.au/pages/p00fset/whatsupf.htm		Educational	
www.lib.mq.edu.au/resources/otherlibraries/	opac.library.usyd.edu.au/screens/opacmenu.html		Technical	
www.newcastle.edu.au/department/el/text%26tech/students /assignt3/melmeth/New%20Folder/hypertextfour.htm	www.ubalt.edu/ygcla/sam/essays/prezones.html		Educational	Educational
www.newcastle.edu.au/services/iesd/learndevelop/resources /online/chat.htm	www.newaol.com/aim/netscape/adb00.html		Technical	
gnu.cs.ntu.edu.au/brave-gnu-world/issue-12.de.html	home.pages.de/~GNU-Pascal/	Personal	Personal	
www.gis.ntu.edu.au/general/links.html	www.unisa.edu.au/gpb/index.htm		Professional	
www.ntu.edu.au/education/csle/research/ebonics/eb6.html	dir.yahoo.com/Society_and_Culture/Cultures_and_Groups /Cultures/American__United_States_/African_American/Ebonics/		Research Oriented	
gnu.cs.ntu.edu.au/software/hurd/debian-gnu-hurd.html	www.debian.org/		Technical	
www.rmit.edu.au/departments/secretariat/search.html	www.dms.dpc.vic.gov.au/l2d/R/ACT01159/index.html		Professional	Professional
www.lib.rmit.edu.au/bidm/bookhm.htm	www.nas.edu/trb/		Professional	
allan.scu.edu.au/LT%26ID.htm	nlu.nl.edu/ace/Resources/Knowles.html		Research Oriented	
www.its.swin.edu.au/systems/stats/proxy/days /domains/proxy.575.domains.html	www.unimelb.edu.au/pwebstats/pwebstats.html		Technical	
www.ld.swin.edu.au/ebusiness/html/links.htm	www.iway.com.au		Professional	
www.swin.edu.au/lib/tutorial/welcome.html	manta.library.colostate.edu/howto/		Navigational	
agbu.une.edu.au/~aaabg/aaadate.html	www.dpie.gov.au/prdc/apsa/		Research Oriented	
abri.une.edu.au/ilrbeefcattle.html	www.hereford.com.au	Administrative	Administrative	
www.une.edu.au/trdc/MTN.HTML	http://www.hreoc.gov.au/disability_rights/index.html[2]		Research Oriented	

lambplan.une.edu.au/breeders/links.htm	www.abrona.com.au/		Personal	
www.its.unimelb.edu.au/ma/public/machttp /machttp-talk-digest-V1-482.html	www.apple.com.au/MPG/		Technical	
www.lib.unimelb.edu.au/collections/medicine/page4.html	cancernet.nci.nih.gov/index.html		Administrative	
www.ecom.unimelb.edu.au/actwww/othersites.html	fisher.stats.uwo.ca/		Administrative	Administrativ
www.ceic.unsw.edu.au/staff/Vicki_Chen/RChan1/rchan1.htm	www.science.uts.edu.au/depts/chem/cteg/		Personal	
newt.phys.unsw.edu.au/~mgb/washpost_990607.html	www.washingtonpost.com/RealMedia/ads/click_lx.ads /www.washingtonpost.com/wp-srv/national/longterm/science/dink/3417/Left/TFK-NAT-4/tobacco234b.gif /38313565303631653337363965313230		Professional	
www.petrol.unsw.edu.au/links/links.html	www.onthenet.com.au/%7Efinke/resource.htm		Other	
anatomy.med.unsw.edu.au/waite/1brain_injury/braininjury.htm	rsb.info.nih.gov/nih-image/Default.html		Research Oriented	
www.uow.edu.au/eng/phys/lecturenotes /modernphysics/modern1/mod131.html	www.terrymorse.com/		Personal	
www.uow.edu.au/arts/histpol/hist361/part4/ghindex.html	glimpse.cs.arizona.edu		Technical	
www-library.uow.edu.au/Subjects/Crea6Res.htm	www.census.gov/	Professional	Professional	
webdev.uow.edu.au/student/careers/services/calendar.html	www.tafensw.edu.au/		Educational	
www.uq.edu.au/~svdpark/Bookmarks.html	www.jokeaday.com/f2babe.htm		Personal	
www.uq.edu.au/~uejchris/auslink.htm	www.travel-library.com/pacific/australia/stybr-language.html		Educational	
www.usq.edu.au/kumbari/index.htm	www.austlii.edu.au/au/other/hreoc/		General/Informative	
www.usq.edu.au/opacs/cllt/sonjb/resources.html	www.aitech.ac.jp/~iteslj/Articles/Kelly-Guidelines.html		Educational	
www.chem.usyd.edu.au/~turner_p/smcf/others.htm	www.amnesty.org.au/		Professional	
www.physics.usyd.edu.au/hienergy/nomad.html	www.pi.infn.it/		Professional	
www-personal.usyd.edu.au/~markz/grevillea/	www.dpie.gov.au/agfor/pbr/pbr.html		Educational	
www.usyd.edu.au/su/course/html/htmlmod5.htm	quest.jpl.nasa.gov/PNG/		Technical	
www.dab.uts.edu.au/ce/research.htm	www.rics.org.uk		Professional	
www.iim.uts.edu.au/learning/confer.shtml	conferences.rpd.net/		Research Oriented	
ftoomsh.progsoc.uts.edu.au/~geldridg/cpp/cppcv3/sect7/	www.cs.rpi.edu/~musser/stl.html		Technical	
www.library.uwa.edu.au/Resources/subject_pages/chemistry.html	www.ccl.net/ccl/acs-fall97/index.html		Research Oriented	Research Oriented
scholar.nepean.uws.edu.au/~n9817666/hindi.html	35.8.242.52/songs/athara.au		Personal	

scholar.nepean.uws.edu.au/~n9815020/japan/	www.winamp.com	Technical	Technical	
www.business.vu.edu.au/bho2250/content_providers.htm	www.ehime-np.co.jp/		Professional	
www.vu.edu.au/careers/profassociation.html	www.iia.net.au/index2.html		Professional	
www.adfa.oz.au/ASEC/	http://idun.itsc.adfa.edu.au/ASEC/		Technical	
www.amc.edu.au/mte	http://www.amc.edu.au		Administrative	
wwwmaths.anu.edu.au/~jurman/	www.encyclopedia.com/browse/34.html		Technical	
complex.csu.edu.au/complex/library/	www.w3.org/vl/Physics/Overview.html		Educational	Educational
www.curtin.edu.au/curtin/dept/ccs/journalism.htm	www.abc.net.au/rn/talks/8.30/mediarpt/mstories/index.htm		Professional	
arts.deakin.edu.au/ir/links/journals.html	www.enews.com/magazines/ajr/		Research Oriented	
som.flinders.edu.au/FUSA/Cochrane/cochrane/ccweb.htm	www.bmjpg.com/data/bkind.htm		Research Oriented	
www.jcu.edu.au/office/research_office/funding /ARC/ARC2001/downloads.html	www.detya.gov.au/highered/research/grants/grantap1.htm		Research Oriented	
www.bendigo.latrobe.edu.au/dssg/web.html	www.telstra.com.au/		Technical	
www.monash.edu.au/serg/docs/linkdocs/uslink.htm	www.pa.utulsa.edu/		Educational	
savanna.ntu.edu.au/publications/proceedings/firesc.html	www.fire.wa.gov.au/		Professional	
www.lib.rmit.edu.au/pathfinders/media-art.html	muse.jhu.edu/		Research Oriented	
www.scu.edu.au/schools/rsm/marine_studies/infoF.html	www.its.csiro.au/		Personal	
www.ld.swin.edu.au/ebusiness/html/links.htm	www.adelaidebank.com.au		Professional	
www.ms.unimelb.edu.au/Links/computing.html	www.cs.ubc.ca/nest/imager/contributions/scharein/KnotPlot.html	Navigational	Research Oriented	Research Oriented
anatomy.med.unsw.edu.au/wwwlink/mac.htm	www.cmpnet.com/search/CMP_Sites:NetGuide		Technical	
www.usq.edu.au/users/klebansk/gen_sci.htm	archives.math.utk.edu/		Educational	
ftoomsh.progsoc.uts.edu.au/~wormwood/darcy/sounds.html	www.adore.net/		Social/leisure	
www.arts.monash.edu.au/others/calico/review/engteach00.htm	.evergreen.loyola.edu/~lmorgan/		Personal	

Appendix 8: Randomly Selected Link Data for UK Universities 2000

Page	Link	Pilot Study	Primary Motivation	Cross-check
blake.sunderland.ac.uk/~ha5tno/links.htm	www-2nd-cs.dcs.st-and.ac.uk/~anb		Personal	
business.dis.strath.ac.uk/guides/	link.bubl.ac.uk/interneteducation		Technical	
cwis.livjm.ac.uk/index/	www.lsh.liv.ac.uk		Educational	
cwis.livjm.ac.uk/lea/info/sci/engineer.htm	www.dis.strath.ac.uk/business/		Professional	
homepages.strath.ac.uk/~cjbs17/computing/systems.html	foldoc.doc.ic.ac.uk/		Technical	
IBS.derby.ac.uk/psychology/links.html	www.sosig.ac.uk/		Personal	
jura1.eee.rgu.ac.uk/cal/resources/cal.html	www.surrey.ac.uk/MechEng/research/CEET/QUEST.html	Research Oriented	Research Oriented	
library.newport.ac.uk/database.htm	edina.ed.ac.uk/art-abstracts/docs/artqrg.pdf		Technical	
library.newport.ac.uk/news.htm	www.niss.ac.uk/news/collections.html		Professional	
libwww.essex.ac.uk/wej4.htm	www.bids.ac.uk		Technical	Technical
lispstat.alcd.soton.ac.uk/dept/links.html	www.lshtm.ac.uk/eps/cps/cpsintro.htm		Professional	
lists.stir.ac.uk/educational-technology-list/latest/msg00013.html	www.staffs.ac.uk/cose		Professional	
psyserver.pc.rhbnc.ac.uk/vision/hoffmann.m/mh/mh_cv.html	www.biols.susx.ac.uk/faculty/biology/russell.htm		Personal	
shelob.iti.salford.ac.uk/~dan/digest/msg00129.html	www.hud.ac.uk/scom/research/Artform/sigplan.html		Research Oriented	
silver.bton.ac.uk/sw/edures.html	www.nhm.ac.uk/		Professional	
sol.brunel.ac.uk/~jarvis/bola/ethics/corporate.html	www.dar.cam.ac.uk/nexus/gamkel.html		Research Oriented	
sun.rhbnc.ac.uk/Classics/resources/links.html	www.bristol.ac.uk/Department/Archaeology /html/homep1.ht		Professional	
web.soi.city.ac.uk/~pw/logsites.html	www.cm.cf.ac.uk/htbin/Graphs/show_stats_graphs		Technical	
www.abdn.ac.uk/logreports/referer.htm	users.ox.ac.uk/~oulsc/competitions/comp.html		Technical	
www.abs.aston.ac.uk/staff/deac/home.htm	www.unl.ac.uk/~carterc/rule.htm		Research Oriented	
www.bath.ac.uk/BUCS/Software/chstdeal.htm	www.chest.ac.uk/software/passport/overview.html		Educational	
www.bio.uea.ac.uk/wwwjournals.html	www.bids.ac.uk		Technical	Technical
www.brad.ac.uk/acad/biomed/FRAMED/INFO/INDINFO.html	www.graylab.ac.uk/		Professional	
www.bris.ac.uk/Depts/English/net_page.html	www.lib.ox.ac.uk/libraries/		Technical	
www.brookes.ac.uk/courses/ugcourses/babsc_joint_bio.html	www.ucas.ac.uk		Educational	
www.brookes.ac.uk/services/library/geol.html	wos.mimas.ac.uk/		Research Oriented	

www.cam.anglia.ac.uk/www/archive_links.html	www.mirror.ac.uk/		Navigational	
www.ch.umist.ac.uk/ramesh.htm	www.soton.ac.uk		Administrative	
www.chem.leeds.ac.uk/WhatsNew.html	ukoln.bath.ac.uk/elib/intro.html		Technical	
www.city.ac.uk/martin/	www.csc.liv.ac.uk/~simon/web/design.html	Technical	Technical	
www.cms.dmu.ac.uk/IRC/sockets.html	www-groups.dcs.st-and.ac.uk/history /Mathematicians/Escher.html		Personal	
www.coventry.ac.uk/stats/daily/dailystats.Apr_08_1999.html	www.statslab.cam.ac.uk/~sret1/analog/		Technical	
www.crg.cs.nott.ac.uk/~dns/dave.html	www.csc.liv.ac.uk/users/team-it/		Personal	
www.cryst.bbk.ac.uk/PPS2/course/section10/small.html	www.bio.cam.ac.uk/scop		Technical	
www.cs.cf.ac.uk/user/C.M.Sully/links.html	http:http:/www.cs.cf.ac.uk/User/C.M.Sully/research.html		Personal	Research Oriented
www.dcs.napier.ac.uk/~dbenyon/family.html	www.le.ac.uk/CWIS/AD/PO/CP/cp.html		Personal	
www.dundee.ac.uk/links/reference/unilinks.htm	www.gcal.ac.uk		Administrative	
www.dur.ac.uk/~dps3em/vision.html	www.cs.cf.ac.uk/Dave/Vision_lecture /Vision_lecture_caller.html		Educational	
www.dur.ac.uk/History/archives.html	www.ncl.ac.uk/library/speccoll/guide.html	Technical	Technical	
www.ex.ac.uk/~SJMacwil/lib/hist.html	www.bodley.ox.ac.uk/boris/guides/maps/mapcase.htm		Professional	
www.ex.ac.uk/~SJMacwil/lib/ssgen.html	www.niss.ac.uk/reference/opacs.html		Technical	
www.geog.le.ac.uk/cti/ggm/cticen.html	www.ph.surrey.ac.uk/cti/home.html		Educational	Educational
www.geog.ucl.ac.uk/~rmunton/bio.htm	www.bham.ac.uk/		Administrative	
www.goldsmiths.ac.uk/libweb/external/weblibs.html	www.warwick.ac.uk/services/library/library.html		Technical	Technical
www.gre.ac.uk/~k.mcmanus/audiovisual.html	www.comlab.ox.ac.uk/archive/images/3d.html		Personal	
www.gre.ac.uk/directory/earthsci/links1.htm	boris.qub.ac.uk/shane/arc/ARChome.html		Personal	
www.herts.ac.uk/lis/subjects/natsci/ejournal/catchwor.htm	www.jstor.ac.uk/journals/00324663.html	Research Oriented	Research Oriented	
www.heythrop.ac.uk/fac/phillinks.html	users.ox.ac.uk/~worc0337/phil_topics.html		Professional	
www.hud.ac.uk/schools/human+health/behavioural_science/staff/AT.html	www.jtap.ac.uk/projects/index.html		Professional	Professional
www.hw.ac.uk/careers/joblinks.html	www.strath.ac.uk/Departments/Careers/guide/		Personal	
www.hw.ac.uk/libWWW/irn/irn28/irn28d.html	www.niss.ac.uk/welcome/whatsnew.html		Personal	
www.ioe.ac.uk/ioe_res/esrc.htm	www.esrc.ac.uk/prog/teaching.htm		Educational	
www.ioe.ac.uk/tcru/sexrep.htm	www.socstats.soton.ac.uk/cshr/safepasseges.htm		Professional	

www.keele.ac.uk/depts/cs/Staff/Homes/Stephen/at.html	link.bubl.ac.uk/computing/		Navigational	
www.keele.ac.uk/depts/cs/Staff/Homes/Stephen/Internet/eghtm1.htm	www.rgu.ac.uk/~sim/research/netlearn/callist.htm		Technical	
www.kingston.ac.uk/~bs_s024/bookmarks.html	sable.ox.ac.uk/ota/		Technical	
www.kingston.ac.uk/~de_s154/3D-Design/Sites_Museums.html	www.vam.ac.uk/		Personal	
www.lamp.ac.uk/library/languages.htm	ahds.ac.uk/		Professional	
www.lboro.ac.uk/computing/services/hallservice/StuReg.html	lanlord.lut.ac.uk/hall.service/		Educational	
www.le.ac.uk/education/resources/more.html	www.bham.ac.uk/education/maths/links/		Educational	
www.lgu.ac.uk/as/library/moorgate/old_stuff/parttime.htm	www.bton.ac.uk/	Administrative	Administrative	
www.lgu.ac.uk/deliberations/subj-rev/aspects.html	www.bbk.ac.uk/Learning/ASD/srh/annexa.html		Educational	
www.liv.ac.uk/Chemistry/Links/infochem.html	mimas.ac.uk/crossfire		Educational	
www.lshtm.ac.uk/eph/mceu/Programme.htm	www.abdn.ac.uk/obsgynae/ogdbc.hti		Research Oriented	Research Oriented
www.materials.qmw.ac.uk/links.html	www.materials.ox.ac.uk/		Research Oriented	
www.ncl.ac.uk/srs/ortho/research.htm	www.dur.ac.uk	Administrative	Administrative	
www.nott.ac.uk/~paxjc/links.htm	www.liv.ac.uk/ctibiol.html		Professional	
www.rcm.ac.uk/research/research/tltp3.html	www.ncteam.ac.uk/		Educational	
www.rdg.ac.uk/AcaDepts/ld/Philos/jmpwebsites.htm	www.liv.ac.uk/Philosophy/depts.html		Professional	
www.salford.ac.uk/ais/publica/subguide/pch.html	bubl.ac.uk		Technical	
www.sbu.ac.uk/ace/workshop/workshop.htm	ace.ac.uk/workshops/archive/index.html		Professional	
www.scit.wlv.ac.uk/~cm1914/java1.2/api/javax/swing/undo /package-summary.html	java.sun.com/docs/books/tutorial/uiswing/components /generaltext.html		Educational	
www.shef.ac.uk/~lib/useful/refs.html	www.bournemouth.ac.uk/library2/html/guide_to_		Educational	Educational
www.shef.ac.uk/uni/projects/vwml/	www.leeds.ac.uk/music/Info/RRTuneBk/tunebook.html		Personal	
www.shu.ac.uk/virtual_campus/cnl/team/cea.htm	www.mailbase.ac.uk/lists/costs-of-networked-learning/	Research Oriented	Research Oriented	
www.soas.ac.uk/Links/uol.html	www.rhbnc.ac.uk/		Administrative	
www.soas.ac.uk/Needham/SCC	www.cup.cam.ac.uk		Educational	
www.soc.soton.ac.uk/OTD/asub/Linkstoothers.html	www.nerc.ac.uk		General/Informative	
www.sos.bangor.ac.uk/madog/index.htm	www.nbi.ac.uk/appl/tglist.html		Technical	Other
www.ssees.ac.uk/british.htm	acdc.hensa.ac.uk/		Technical	
www.ssees.ac.uk/july99.htm	www.cspp.strath.ac.uk/		Research Oriented	

www.surrey.ac.uk/Library/er.html	www.niss.ac.uk/lis/obi/obi.html		Technical	
www.tvu.ac.uk/uniservices/content/tvustudentstaff/web/subscription.html	www.bids.ac.uk/education		Technical	
www.uce.ac.uk/study_ops/lss/courses/certsocial.htm	www.sbirmc.ac.uk/		Navigational	
www.ucl.ac.uk/Resources/Searching/findmail.htm	www.mailbase.ac.uk/	Personal	Personal	
www.uclan.ac.uk/library/libelj3.htm	pinkerton.bham.ac.uk/rpsv/catchword/carfax/13569317 /contp1-1.htm		Research Oriented	
www.uclan.ac.uk/prospectus/courses/partner/files/burndata.htm	www.burnley.ac.uk/		Administrative	
www.uel.ac.uk/sociology/euro/socrates/page9.html	www.hefce.ac.uk		Professional	
www.ulh.ac.uk/ls/staff/jeb/research.htm	www.bids.ac.uk		Technical	
www.ulst.ac.uk/cticomp/english.html	computing.unn.ac.uk/	Professional	Professional	
www.unl.ac.uk/library/ess/humageog.shtml	wos.mimas.ac.uk		Research Oriented	
www.uwcm.ac.uk/uwcm/pr/links/nursing.html	www.csv.warwick.ac.uk:8000/resources/nurse-resources/		Professional	
www.uwe.ac.uk/library/resources/general/electronic_journals/jnlsa_b.htm	www.ariadne.ac.uk		Research Oriented	
www.wye.ac.uk/AtoZ/index_b.html	www.uclan.ac.uk/biotutor		Professional	
www.york.ac.uk/inst/ctipsych/web/CTI/DirTxt/reviews/macretina.html	www.cti.ac.uk/centres		Navigational	
www2.wales.ac.uk/non-frames/english/Study/awards.html	www.cf.ac.uk		Administrative	Administrat
www2.wmin.ac.uk/library/libfaq.html	www.bids.ac.uk		Technical	
www-icdl.open.ac.uk/icdl/export/europe/unitedki/fifecoll/inst/index.htm	www.fife.ac.uk	Administrative	Administrative	
www-scm.tees.ac.uk/internet/unis.html	www.bangor.ac.uk		Administrative	
www-tec.open.ac.uk/eeru/staff/scthm/mmo/leaf.html	www.crg.cs.nott.ac.uk/ukvrsig/conf/sigconf-1.html		Professional	

Appendix 9: Randomly Selected Link Data for New Zealand Universities 2006

Page	Link	Pilot Study	Primary Motivation	Cross-check
www.education.auckland.ac.nz/learning/science/links/earth.asp	socrates.berkeley.edu/~eps2		Educational	
www.trout.auckland.ac.nz/journal/12/12_18.html	www.unesco.org		Administrative	
www.math.auckland.ac.nz/~waldron/Hermite/hermite.html	pauillac.inria.fr/algo/bsolve/constant/ws/ws.html		Research Oriented	
www.library.auckland.ac.nz/subjects/bus/course-pages /mktg306_marketing_comm.htm	search.epnet.com/direct.asp?an=3712818&db=buh		Research Oriented	
www.esc.auckland.ac.nz/people/staff/mehr002/Reference s/Author/HATCHUEL-A.html	www-sop.inria.fr/epidaure/personnel/malandain/codes /bibtex2html.html		Technical	
transitofvenus.auckland.ac.nz/	www.natlib.govt.nz		Superficial	
www.library.auckland.ac.nz/subjects/bio/course-pages/biosci_206.htm	search.eb.com/	Educational	Educational	
www.math.auckland.ac.nz/Conferences/TIME2000/AuckNZ.html	www.immigration.govt.nz/		Administrative	
www.library.auckland.ac.nz/subjects/asian/china.htm	art-design.umich.edu/mother/resource.html		Research Oriented	
www.library.auckland.ac.nz/subjects/pol/course-pages/politics222_2.htm	search.epnet.com/login.aspx?direct=true&db= aph&an=8962500		Technical	
www.library.auckland.ac.nz/subjects/linguistics/lingsresguide.htm	anthro.AnnualReviews.org/cgi/content/full/26/1/47		Research Oriented	
www.math.auckland.ac.nz/PhD/	www.math.waikato.ac.nz/NZMS/support.html		Professional	Professiona
www.nzcel.auckland.ac.nz/comparative.htm	www.environment.sa.gov.au/		Professional	
www.library.auckland.ac.nz/subjects/anthro/course-pages/105300.htm	www.vlib.org/		Technical	
www.library.auckland.ac.nz/subjects/eng/resources/netcm.htm	www.memsnet.org/material/		Research Oriented	
www.comsdev.canterbury.ac.nz/diary/2002/020426.htm	www.uni-care.org/		Professional	Professiona
www.cosc.canterbury.ac.nz/~greg/yacc++/yacc++.1.html	localhost/usr/local/man?lex?1		Technical	
library.canterbury.ac.nz/law/subjects/tax.shtml	www.ccra-adrc.gc.ca/		Educational	
www2.phys.canterbury.ac.nz/kerrfest/travel.html	www.qantas.com.au/regions/dyn/homeInternationalSearc h		Social/leisure	
www.soci.canterbury.ac.nz/resources/glossary/unintend.shtml	www.comminit.com/changetheories/ctheories /changetheories-63.html		Educational	
library.canterbury.ac.nz/collserv/bookseln.shtml	www2.h-net.msu.edu/reviews/		Professional	
www.ucar.canterbury.ac.nz/research/activelks.htm	www.geo.vuw.ac.nz/meteorology/	Research Oriented	Research Oriented	
library.canterbury.ac.nz/educ/web/child_dev.shtml	www.brainwave.org.nz		Research Oriented	
www.canterbury.ac.nz/student/careers/links/academic.shtml	linguistlist.org		Educational	

www.math.canterbury.ac.nz/~mathmas/places.html	www.mis.mpg.de/conferences/phylo2004/		Research Oriented	
www.econ.canterbury.ac.nz/links/oseso.shtml	cepr.anu.edu.au/		Administrative	
library.canterbury.ac.nz/web/eref/libs.shtml	www.shorelibraries.govt.nz		Technical	Technical
www.canterbury.ac.nz/student/careers/subjectoptions/tafs.shtml	www.prospects.ac.uk		Educational	
library.canterbury.ac.nz/web/eref/intergovtorgs.shtml	www.nato.int/		Professional	
www.lincoln.ac.nz/libr/natres.htm	www.ncedr.org/tools/othertools/costbenefit/lead.htm	Administrative	Administrative	
www.lincoln.ac.nz/libr/educ.htm	www.studygs.net		Educational	
www.lincoln.ac.nz/libr/genref.htm	www.skepdic.com		Social/leisure	
www.lincoln.ac.nz/libr/libcats.htm	nlnzcat.natlib.govt.nz	Technical	Technical	
events.lincoln.ac.nz/apbc/partner.htm	www.molecularplantbreeding.com		Superficial	
www.lincoln.ac.nz/libr/dbases/dbalpha.htm	library.wur.nl/desktop/catalog/		Technical	
www.lincoln.ac.nz/libr/dbases/access.htm	www.ajol.info		Research Oriented	
events.lincoln.ac.nz/treasury/	www.kpmg.co.nz		Superficial	
www.lincoln.ac.nz/libr/news/newsamer.htm	www.emol.com/		Professional	
www.lincoln.ac.nz/libr/subjects/maoriind.htm	www.lawsite.co.nz/landcare/		Educational	
www.lincoln.ac.nz/libr/genref.htm	www.lib.umich.edu/govdocs/stats.html		Professional	
www.lincoln.ac.nz/libr/dbases/maptoast.htm	www.linz.govt.nz/rcs/linz/pub/web/root/core/Topography /aerialandorthophotos/index.jsp		Technical	
www.lincoln.ac.nz/libr/news/newseuro.htm	www.kurier.at		Technical	Other
events.lincoln.ac.nz/apbc/general.htm	www.agresearch.co.nz/		Superficial	
www.lincoln.ac.nz/libr/dbases/wos.htm	www.isinet.com/tutorials/wos6/		Research Oriented	
www.lincoln.ac.nz/libr/dbases/dbcomp.htm	firstsearch.oclc.org/fsip		Technical	Technical
www.massey.ac.nz/~wwmansys/staff/walker_john.htm	www.mymassey.com	Personal	Personal	
www.massey.ac.nz/~tameyer/writing/insecure_comments.html	validator.w3.org/check/referer		Superficial	
masseynews.massey.ac.nz/2005/Press_Releases/12-21-05.html	www.pncc.govt.nz/	Administrative	Administrative	
misg2005.massey.ac.nz/goals.html	www.maths.monash.edu.au/anziam06		Research Oriented	
masseynews.massey.ac.nz/2004/Press_Releases/07_30_04.html	www.northshorecity.govt.nz/		Administrative	
masseynews.massey.ac.nz/2003/masseynews/feb/feb24/benchmarking.ht ml	www.pncc.govt.nz/		Professional	
research.massey.ac.nz/commercialisation/com_academic.htm	www.mymassey.com		Personal	

www.massey.ac.nz/~ychisti/FileF.html	www.springeronline.com/sgw/cda/frontpage/0,11855,1-0-70-35661427-detailsPage%253Djournal%257CeditorialBoard%257CeditorialBoard,00.html?referer=www.springeronline.com/journal/10811/edboard		Research Oriented	
www.massey.ac.nz/~chmessom/159339/index.html	www.comptechdoc.org/independent/web/http/reference/index.html	Professional	Professional	
www.massey.ac.nz/%7Eimbs/HTML/penny.html	awcmee.massey.ac.nz/people/dpenny/abstract_evol.htm		Research Oriented	
praxis.massey.ac.nz/prism_on-line_journ.html	www.bond.edu.au/hss/		Research Oriented	Research Oriented
education.massey.ac.nz/ICT-Online/teacherpd.asp	www.unitec.ac.nz/education/progs/index.html		Educational	
www.business.otago.ac.nz/Sirc/webpages/Conferences/SIRC2002/index.html	www.geohealth.org.nz/		Professional	
www.otago.ac.nz/NZCMSWebTable/VICSABEATSRR.html	www.wnmeds.ac.nz/Infoserv/index.html		Navigational	
www.business.otago.ac.nz/finc/links.html	www.dfin.com		Professional	
www.business.otago.ac.nz/tourism/links/index.html	www.trcnz.govt.nz		Research Oriented	
www.cs.otago.ac.nz/postgrads/alexis/links.html	www.wetafx.co.nz/		Personal	
www.otago.ac.nz/NZCMSWebTable/VDCJABEATSRR.html	www.wnmeds.ac.nz/academic/index.html		Navigational	
www.library.otago.ac.nz/subject_guides/region.html	www.colorado.edu/geography/virtdept/resources/journal/journals.htm		Research Oriented	
www.covic.otago.ac.nz/~ddeng/	www.inns.org		General/Informative	
www.physics.otago.ac.nz/versim/versim07.html	www.oulu.fi/~spaceweb/research.html		Research Oriented	
www.cs.otago.ac.nz/staffpriv/ok/HOS2002/Luke-Hamilton.html	hotwired.lycos.com/webmonkey/	Technical	Technical	
secml.otago.ac.nz/dcsa/index.html	home.netscape.com/download/index.html?cp=djuc1		Superficial	
www.otago.ac.nz/NZCMSWebTable/VICPAGEATSRR.html	www.wnmeds.ac.nz/index.html		Professional	
www.otago.ac.nz/NZCMSWebTable/VICJAFEATSR.html	www.wnmeds.ac.nz/academic/dph/research/index.html		Research Oriented	
www.otago.ac.nz/NZCMSWebTable/VECBAHEATSRD.html	www.wnmeds.ac.nz/academic/dph/index.html		Superficial	Superficial
www.vuw.ac.nz/~caplabtb/dprk/DPRK_economy.html	www.kcna.co.jp/item/2001/200103/news03/21.htm		Educational	
www.vuw.ac.nz/~caplabtb/dprk/NK_paper.htm	www.japantimes.co.jp/cgi-bin/geted.pl5?eo20030206lf.htm		Research Oriented	
www.vuw.ac.nz/~caplabtb/dprk/NK_US_00.htm	www.koreaherald.co.kr/news/2000/09/__02/20000923_0209.htm		Professional	
www.vuw.ac.nz/~caplabtb/dprk/NK_misc.htm	joongangdaily.joins.com/200411/19/20		Social/leisure	
www.mcs.vuw.ac.nz/courses/COMP423/2005T1/project1	www.asx.com.au/clickable/change.htm	Social/leisure	Social/leisure	

/Prices%20Results_5.htm				
www.vuw.ac.nz/~caplabtb/dprk/NK_tour.htm	english.joins.com/nk/news/Media/Img/golf.jpg		Social/leisure	Social/leisu
www.vuw.ac.nz/~caplabtb/dprk/SK_0506.htm	joongangdaily.joins.com/200506/24/20		Social/leisure	
www.vuw.ac.nz/~caplabtb/dprk/peninsula_may02.html	english.joins.com/nk/article.asp?aid= 20020517102018&sid=E00		Professional	
www.utdc.vuw.ac.nz/blackboard/technical.shtml	www.apple.com/quicktime/download/index.html		Technical	
www.vuw.ac.nz/~caplabtb/dprk/DPRK_russia.html	www.atimes.com/c-asia/CB23Ag01.html		Professional	
www.vuw.ac.nz/art-history/courses/arth218.html	www.adobe.com/products/acrobat/readstep2.html		Technical	Technical
www.mcs.vuw.ac.nz/~visser/pictures.shtml	www-spires.slac.stanford.edu/find/hep		Research Oriented	
www.vuw.ac.nz/~caplabtb/dprk/NK_china.htm	www.washingtonpost.com/wp-dyn/content/article /2005/10/21/AR2005102101099.html		Professional	
www.vuw.ac.nz/~caplabtb/dprk/NK_gen.htm	times.hankooki.com/lpage/200512 /kt2005121817233510220.htm		Professional	
www.vuw.ac.nz/~caplabtb/dprk/DPRK_miscellaneous.html	www.marathonguide.com/news/newsviewer.cfm?src=		Professional	
www.waikato.ac.nz/library/resources/law/s_uk.shtml	www.publications.parliament.uk/pa/ld/ldjudinf.htm		Educational	
webteam.waikato.ac.nz/caudit-report/	www.caudit.edu.au/		Administrative	
www.waikato.ac.nz/wfass/subjects/geography/papers/03300 /examples/heritage-hamilton/Page%207.htm	www.zoo.org.au		Social/leisure	
www.cs.waikato.ac.nz/~masood/teaching/224a/links.html	www.synfonts.com/		Educational	
www.waikato.ac.nz/film/student/2001/0211310B/JennyChiu/Links.html	sun3.lib.uci.edu/~dtsang/arc.htm		Research Oriented	
www.waikato.ac.nz/sasd/careers/links.shtml	www.thedailynews.co.nz		Professional	
www.cs.waikato.ac.nz/~marku/soundcards.html	www.linuxlinks.com/Software/System/DeviceDrivers/		Technical	
www2.waikato.ac.nz/library/resources/subject_portal/fass_arts.shtml	webdirectory.natlib.govt.nz/dir/en/nz/arts-and-literature /architecture/		Social/leisure	
www2.waikato.ac.nz/library/learning/s_mnfp.shtml	en.wikipedia.org/wiki/Donations		Social/leisure	
www.cs.waikato.ac.nz/TNG/middlemiss.html	www.acm.org		Administrative	Administrat
www.waikato.ac.nz/library/resources/edu/science_environed.shtml	www.eednz.org.nz		Technical	
www.cs.waikato.ac.nz/~jcleary/230/TIJ/html/Chap14.htm	www.mindview.net/backtalk/CommentServlet? ID=TIJ3_CHAPTER14_I99		Other	
www.vuw.ac.nz/home/index.asp	www.vuw.ac.nz/home/research/overview.html	Research Oriented	Research Oriented	
www.canterbury.ac.nz/courses/	.canterbury.ac.nz/help/legal.shtml		Educational	

Appendix 10: Randomly Selected Link Data for Australian Universities 2006

Page	Link	Pilot Study	Primary Motivation	Cross-c
www.cecs.acu.edu.au/cyprian/othersites.php	www.ccel.org/fathers2		Professional	
dlibrary.acu.edu.au/databases/dbListE.asp	bondi.unilinc.edu.au/acuer		Research Oriented	
dlibrary.acu.edu.au/research/theology/ejournal/aejt_2/Curran.htm	www.ornl.gov/sci/techresources/Human_Genome/home.shtml	Research Oriented	Research Oriented	
dlibrary.acu.edu.au/research/CarpeDiem/pages/journal1/essay1.htm	www.aar.com.au/corpgov/pubs/cdlaug03.htm		Research Oriented	
www.adelaide.edu.au/library/guide/crime/	www.kirjasto.sci.fi/aamilne.htm		Personal	
www.ees.adelaide.edu.au/research/geology/cerg/pubs /pub_gheinson.html	www.agu.org/pubs/crossref/2005/2005GL022934.shtml		Research Oriented	
ajax.acue.adelaide.edu.au/web/Xtras/Flash2Asset /Help/html/FlashHelp.html	www.macromedia.com/support/xtras_essentials/		Technical	
www.adelaide.edu.au/library/guide/gen/essay/	www.unisanet.unisa.edu.au/learningconnection/students /lrnsvcs/refncing.asp		Educational	Educati
wwwrsphysse.anu.edu.au/~ask107/publications.html	www.iop.org/EJ/S/3/67/RjpXf68Bcq7oTo.Qyv3Ebg /abstract/0953-8984/11/35/314		Research Oriented	
brf.jcs.anu.edu.au/services/fluophosimager/Fluoimager.html	home.fujifilm.com/products/science/application/index.html		Navigational	
www.anu.edu.au/people/Roger.Clarke/EC/FDST.html	www.discover.com/feb_01/gthere.html?article=featnapster.html		Research Oriented	
uob-community.ballarat.edu.au/live/mar04/fast_facts.htm	www.ballarat.net.au/big/ratnet_live.ram		Educational	Other
james.bond.edu.au/shared/docbook/tip.html	www.oreilly.com/		Technical	
www.it.bond.edu.au/inft132/INFT132Site/132Links.htm	www.awrestaurants.com		Social/leisure	
uctv.canberra.edu.au/Members/geo/new-date-test/sendto_form	plone.org		Technical	
www.blis.canberra.edu.au/lawportal/more/topics.htm	www.dfat.gov.au/hr/		Educational	
www.canberra.edu.au/maygibbs/links.html	www.maygibbs.org.au/		Personal	
www.blis.canberra.edu.au/schools/law/resources/links.htm	jurist.law.pitt.edu/subj_gd.htm		Research Oriented	
www.ntu.edu.au/education/oll/collab/activities.html	138.80.22.25/bin/common/control_panel.pl?course_id=_49_1	Educational	Educational	
informatics.ntu.edu.au/staff/kgilbert/hit362/index.html	www.tux.org/lkml/		Technical	
clp.cqu.edu.au/online_articles.htm	www.press.umich.edu/jep/06-01/payne.html		Research Oriented	
ahe.cqu.edu.au/other_sites.htm	buros.unl.edu/buros/jsp/search.jsp		Educational	
clp.cqu.edu.au/miscellanea.htm	www.online-seminar.net		Technical	
www.engineering.cqu.edu.au/engcoop/employerlist/elecemployers.htm	www.qal.com.au/		Professional	Professi

hsc.csu.edu.au/hospitality/hosp_240/accom_serv /THTSOP06B/reservations/THHBFO01AMP.html	www.boardofstudies.nsw.edu.au/hsc_exams/		Navigational	
hsc.csu.edu.au/cafs/resources/2274/index.htm	www.schools.nsw.edu.au/		Administrative	
learnt.smec.curtin.edu.au/resources/classroom_resources /ontario/maths/gr7nsan.html	www.funbrain.com/cracker/index.html		Social/leisure	
lsn.curtin.edu.au/tlf/tlf2002/ho.html	www.aare.edu.au/99pap/bre99209.htm		Research Oriented	
research.curtin.edu.au/grants/grantadmin.html	www.dest.gov.au/sectors/research_sector/programmes_funding /general_funding/research_infrastructure /research_infrastructure_block_grants_scheme.htm		Research Oriented	
research.wasm.curtin.edu.au/weblinks.php	www.ism.rwth-aachen.de/		Administrative	
www.deakin.edu.au/buslaw/law_research/subject/employment.php	www.acirrt.com/		Professional	
www.deakin.edu.au/buslaw/aef/links/index.php	www.cba.ufl.edu/fsoa/schoolinfo/jal.html		Research Oriented	Researc Oriente
www.deakin.edu.au/dusa/stud_support/postgraduate_financial.php	www.jason.unimelb.edu.au/		Technical	
www.deakin.edu.au/buslaw/law_research/subject/compcases.php	www.austlii.edu.au/au/legis/cth/consol_reg/tpr1974258/		Professional	
www.lib.flinders.edu.au/resources/ej/a-z/a.shtml	ejournals.ebsco.com/direct.asp?JournalID=101072		Research Oriented	
www.nisu.flinders.edu.au/pubs/bulletin9/b9t4.html	www.health.gov.au		Personal	
som.flinders.edu.au/FUSA/ClinPharm/UGT/udgpa.html	www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db= protein&val=31324698		Professional	
www.gu.edu.au/vc/ate/content_links.html	www.crc.gov.au/		Research Oriented	
www.gu.edu.au/school/hsv/content/assistance/hsr /assistance_hsr_internet_resources.html	www.griffith.edu.au/ins/collections/electronic/		Technical	
www.cit.gu.edu.au/teaching/2507CIT/Resources /jwstutorial13/doc/JSPTags.html	java.sun.com/webservices/docs/1.3/tutorial/information /sendusmail.html		Personal	
www.jcu.edu.au/office/itr/caudit_survey2003/Questions/Q9.htm	www.uq.edu.au		Administrative	
www.library.jcu.edu.au/Resources/ej10.shtml	www.blackwell-synergy.com/servlet/useragent? func=showIssues&code=jopo	Research Oriented	Research Oriented	
www.library.jcu.edu.au/Resources/ejfull.shtml	www.cap.org/html/publications/captoday.html		Administrative	
library.bendigo.latrobe.edu.au/irs/webcat/820.htm	www.Suite101.com/links.cfm/teen_issues_book_reviews		Research Oriented	
library.bendigo.latrobe.edu.au/irs/webcat/350.htm	en.wikipedia.org/wiki/Category:Government_of_Jordan		General/Informati ve	Other
www.law.monash.edu.au/postgraduate/master-ic-law.html	www.monash.edu/pubs/handbooks/units/LAW7212.html	Educational	Educational	
www.ctie.monash.edu.au/hargrave/legacy1.html	www.memagazine.org/contents/current/features /palmsize/palmsize.html		Research Oriented	

www.ics.mq.edu.au/~kelp/Corpus/AnnotatedFiles/output0064.html	welcome.hp.com/country/us/eng/solutions.htm		Superficial	Superfic
www.es.mq.edu.au/GEMOC/glitter/glossy.html	www.glitter-gemoc.com/		Administrative	
www.library.mq.edu.au/researchguides/environment/websites.html	www.greenhouse.gov.au		Research Oriented	
wwwlib.murdoch.edu.au/electref/unis.html	www.arc.gov.au/default.htm		Professional	
wwwlib.murdoch.edu.au/guides/arts/brithist.html	sb1.abc-clio.com:81/		Professional	
www.newcastle.edu.au/research/brf/facs/vantage.html	www.tunra.com/		Professional	
www.newcastle.edu.au/faculty/business-law/research/index.html	cricos.detya.gov.au/asp/InstitutionDetails.asp?PVID=109	Administrati ve	Administrative	
www.library.qut.edu.au/subjectpath/chemistry_general.jsp	www.chemsoc.org/		Technical	
www.research.qut.edu.au/oresearch/onlineinfor/qutresearchn /News2005/October%202005/ResNewsIT.jsp	www.arc.gov.au/		Research Oriented	
www.bus.qut.edu.au/research/cpns/whatweresearch/usefullinks.jsp	www.auscharity.org/		Professional	
hypertext.rmit.edu.au/vog/vlog/links/newmedialist.html	www.noisebetweenstations.com/		Personal	
pgconf.cs.rmit.edu.au/Downloads/photos/photo_41.html	www.ornj.net/	Superficial	Superficial	
buildlca.rmit.edu.au/decisiontool/S3materials.html	sun1.mpce.stu.mmu.ac.uk/pages/projects/dfe/deeds/ecodnavi /toolbox/analyse/lcabased/lcinvent/boustead.html		Personal	
www.scu.edu.au/schools/edu/student_pages/sem1_2002 /dmcpherson/ocean.htm	home.mira.net/~areadman/shark.htm		Social/leisure	
chlib.scu.edu.au/management.htm	bubl.ac.uk/link/m/managementresearch.htm		Professional	
php.it.swin.edu.au/ChangeLog-5.php	bugs.php.net/31341		Technical	
mercury.it.swin.edu.au/swinbrain/index.php /Swinburne_Java_Coding_Standard	www.google.com/search?q=allinurl% 3AMouseEvent+java.sun.com&bntl=1		Educational	
www.swin.edu.au/lib/infogate/vocedtraining.htm	www.aqf.edu.au/implem.htm		Educational	
www.une.edu.au/library/elecres/lawlinks.htm	law.anu.edu.au/colin	Professional	Professional	
www.une.edu.au/arts/SouthAsiaNet/sites.htm	www.pathfinder.com/@@ISdtF1GnMwMAQGM8/Asiaweek/		Professional	
www-personal.une.edu.au/~aschalle/cv.html	www.dgfs.de/cgi-bin/dgfs.pl		Administrative	
www.bme.unimelb.edu.au/about/links.html	www.aimbe.org		Administrative	
www.asor.ms.unimelb.edu.au/apors97/program/web309.html	volans.cbr.dit.csiro.au:8001/apors/abstracts/abstract-467.html		Research Oriented	Researc Oriente
www.library.unisa.edu.au/resources/subject/naturop.asp	www.rxlist.com/alternative.htm		Professional	
www.unisa.edu.au/policies/policies/corporate/C23.asp	www.atn.edu.au/		Administrative	Admini
www.unisa.edu.au/future/events.asp	webstandards.org/act/campaign/buc/		Superficial	

cellbiology.med.unsw.edu.au/units/medicine/REbone04.htm	pubs.niaaa.nih.gov/publications/arh25-4/276-281.htm		Research Oriented	
www.library.unsw.edu.au/~libadmin/infolit.html	www.ala.org/acrl/ilcomstan.html		Educational	
www.uow.edu.au/arts/sts/bmartin/pubs/04nexus.htm	www2.coloradocollege.edu/Library/Course /downloading_detectives_paper.htm		Research Oriented	
src.uow.edu.au/hecs_help.php	en.wikipedia.org/wiki/Tertiary_education		Professional	
www.uow.edu.au/research/rso/publications/2002/education.html	www.capstrans.edu.au		Educational	
www.library.uq.edu.au/schools/uq_subscription.html	cricos.detya.gov.au/asp/InstitutionDetails.asp?PVID=025		General/Informati ve	
www.uqunion.uq.edu.au/businessservices/food/default_html	www.schonell.com		Social/leisure	
www.accs.uq.edu.au/index/location	www.ourbrisbane.com/		Educational	
student.uq.edu.au/~s4057040/	cricos.detya.gov.au/asp/InstitutionDetails.asp?PVID=025		Superficial	
www.usc.edu.au/University/Library/Resources/Databases/ A-Z/ProQuest.htm	isi01.isiknowledge.com/portal.cgi/		Research Oriented	
www.usq.edu.au/library/faculties/business/facinternet /subjectsites/mgmt/mobhrm.htm	www.stern.nyu.edu/~wstarbuc/		Personal	
www.usq.edu.au/course/specification/2005 /THS3003-S2-2005-45415.html	www.adobe.com	Technical	Technical	
www.aeromech.usyd.edu.au/~mcbain/computing/	validator.w3.org/check/referer		Technical	
vein.library.usyd.edu.au/links/pact/rd026.html	www.pgf.edu.au/refdb/index.cfm	Professional	Professional	
www.maths.usyd.edu.au/u/AusCat/abstracts/980325sc.html	www.ics.mq.edu.au/CoACT/		Administrative	Adminis
www.cs.usyd.edu.au/~loki/info/email/msn/addr009.htm	www.hotmail.com/cgi-bin/start?login=jordan.dell&passwd= DoiCjBY0Aa&enter=Enter&frames=no&curmbox=ACTIVE&js=no		General/Informati ve	
www.utas.edu.au/mgmt/GSM/delivery.html	www.anbs.com.au		Educational	
oak.arch.utas.edu.au/glossary/glossary20.html	www.tastimber.tas.gov.au		Superficial	
services.eng.uts.edu.au/userpages/johnr/public%5Fhtml/index.html	www.aiki.com.au/		Social/leisure	
www.education.uts.edu.au/japanese/gdlt_faq_japan.html	138.25.124.210/scm/ntbc?f=EDU		Research Oriented	
www.csd.uwa.edu.au/iced2002/abstract/Nunan.html	www.webstandards.org/upgrade/		Technical	
robotics.ee.uwa.edu.au/eyebot/examples-rob/SoccerBot /cluster/src/html/classLCD.html	www.doxygen.org/index.html		Educational	
www.publishing.uwa.edu.au/annualreport/2001 /fi-03-certification_of_2001_financial_statements.asp	jigsaw.w3.org/css-validator/	Superficial	Superficial	
www.uws.edu.au/about/adminorg/devint/ors /destpubscollection/catcodes/journal	www.dest.gov.au/highered/research/herdc.htm		Research Oriented	

w2.vu.edu.au/library/infolink/education/education2.htm	www.diversityweb.org/		Professional	Professi
www.staff.vu.edu.au/CulturalCalendar/Months/05November.html	www.unesco.org/tolerance/teneng.htm		Professional	
toolbox.tafe.vu.edu.au/telecommunications/om/om_c01.html	www.ccma.asn.au/		Professional	

Appendix 11: Randomly Selected Link Data for UK Universities 2005

Page	Link	Pilot Study	Primary Motivation	Cross-c
online.northumbria.ac.uk/csru/links.htm	www.go-ne.gov.uk		Educational	
online.northumbria.ac.uk/geography_research/radix/humanrights5.htm	www.unglobalcompact.org		Research Oriented	
www.abdn.ac.uk/~src119/links.hti	www.glasgowstudent.net/main/studentorganisations /volunteering/dirty_weekenders		Social/leisure	
www.irs.aber.ac.uk/als/netpage/lectures/peopleohp.htm	www.countryside.gov.uk/research/f_state.htm		Educational	
users.aber.ac.uk/smg/WORK/refrnce.htm	scitsc.wlv.ac.uk/ukinfo/uk.map.html		Technical	
www.arts.ac.uk/library/4790.htm	www.artscope.org.uk/		Educational	
www.arts.ac.uk/student/counselling/2063.htm	www.al-anonuk.org.uk/		Educational	
www.aston.ac.uk/lis/subjects/abs/ECommerce.jsp	reports.mintel.com/sinatra/mintel/search/		Technical	
www.aston.ac.uk/lss/staff/profile/gvmonographs.jsp	www.holtmann-mares.de/Varouxakis.htm		Research Oriented	
biology.bangor.ac.uk/research/publication/A1994MR34900005	validator.w3.org/check/referer		Superficial	
staff.bath.ac.uk/ensdgg/bookmarks.htm	www.tiscali.co.uk/reference/dictionaries/difficultwords/		Personal	Other
www.bath.ac.uk/health/news/current.html	www.rdinfo.org.uk/Queries/WhatsNew.asp		Professional	Professi
www.bathspa.ac.uk/departments/student-support/careers /useful-web-links/subject.asp	www.poptel.org.uk/Labour-Party/		Professional	
www.bbk.ac.uk/english/ac/19Cwomensyll.htm	www.indiana.edu/~letrs/vwwp/		Research Oriented	
www.iel.bham.ac.uk/lawyers.htm	/errormessages/404error.htm?404; http://www.iel.bham.ac.uk/lawyers.htm		General/Informative	
www.universitas21.bham.ac.uk/staff/opportunities.htm	www.solander.lu.se/		Research Oriented	
www.student.brad.ac.uk/aalbabah/	www.geeks404.com	Educational	Educational	
www.bio.bris.ac.uk/research/morlab/BiolBull.htm	www.biolbull.org		Research Oriented	
www.bris.ac.uk/Depts/Union/Diving/OtherResources/unidive.html	www.luusac.co.uk/		Social/leisure	
www.brookes.ac.uk/other/unison/HS%20Bulletin%20_0205.html	www.hazards.org/smoking		General/Informative	
people.brunel.ac.uk/~careers/students_jobhunting_A-Z.shtml	www.ukjobsguide.co.uk		Professional	
www.it.bton.ac.uk/research/seake/links.html	www.ids.ac.uk		Administrative	
www.cam.ac.uk/societies/round/band/index/ho-hz.htm	www1.roke.co.uk/SIB/abc/midi/J2_3.mid		Social/leisure	
www.cardiff.ac.uk/carbs/lom/lsdg/euronil/Links/links5.html	members.surfeu.fi/otaniemi/plrpubl.htm		Research Oriented	
www.astro.cardiff.ac.uk/wusage/week137.html	siva.cshl.org/wusage.html		Other	Other

www.chichester.ac.uk/international/index.htm	www.gksoft.com/govt/en/sa.html		Navigational	
legacywww.coventry.ac.uk/legacy/coventry/links.htm	www.cwn.org.uk/tourism/eating-out/index.html		Social/leisure	
www.derby.ac.uk/sehs/ehs-busliaison.asp	validator.w3.org/check/referer		Superficial	
www.dundee.ac.uk/pharmacology/Deadline_Calendar/NoDeadline.htm	www.fanconi.org/grants.html		Research Oriented	
www.dur.ac.uk/r.j.coe/resmeths/litsearch.htm	firstsearch.oclc.org/route=UK;done=referer;FSIP		Research Oriented	
www.essex.ac.uk/freshers/important/missedadvice(ug).htm	www.slc.co.uk/		Professional	Professi
www.ex.ac.uk/german/media/wirtch.html	www.handelszeitung.ch/zeitschriften/logistik/welcome.html		Professional	
www.lib.gla.ac.uk/Subject/History/index.shtml	www.rhd.uit.no/indexeng.html		Research Oriented	
www.comp.glam.ac.uk/pages/staff/dwfarthi/projman.htm	www.bppm.com		Professional	
www.gold.ac.uk/tmr/	www.cordis.lu/tmr/src/network1.htm		Professional	
www.gre.ac.uk/lib/subjects/health/websites.html	www.gingerbread.org.uk/		Educational	
www.harper-adams.ac.uk/leisure/professional_bodies.htm	www.toursoc.org.uk/		Professional	
www.health-homerton.ac.uk/learning/library/bibdb.html	www.bookfind-online.com/		Technical	
www.herts.ac.uk/lis/subjects/natsci/ejournal/ejnllistB.htm	www.sciencedirect.com/science/journal/10773150		Research Oriented	
www.hud.ac.uk/encore/yournews/1980.html	www.walkingholidays.org.uk		Social/leisure	Social/l
www.iconex.hull.ac.uk/links2.htm	150.237.4.182/xtensis/xtguest.htm		Navigational	
www.hw.ac.uk/library/njindexs.html	www.sciencedirect.com/science/journal/0038092X		Research Oriented	
amber.ch.ic.ac.uk/	amber.scripps.edu/		Administrative	
www.kcl.ac.uk/depsta/law/research/coroners/caselinks.html	www.austlii.edu.au/au/cases/vic/VICSC/unrep2.html		Educational	
www.keele.ac.uk/depts/li/hl/nhsjnls/journalv.htm	www.adobe.com/products/acrobat/readstep2.html		Technical	Technic
www.kent.ac.uk/careers/managementscience.htm	www.orsoc.org.uk/about/career/brochure/book.htm		Educational	
www.lamp.ac.uk/cis/liminal/virtuallyislamic/surfingislam.html	saudi.sexypage.net/	Research Oriented	Research Oriented	
the-stable.lancs.ac.uk/~esarie/guestbooks/64guest.htm	www.hometown.aol.com/whassupjustin /myhomepage/pet.html	Personal	Personal	
www.lboro.ac.uk/admin/ar/international/general/contact/index.htm	jigsaw.w3.org/css-validator/check/referer		Superficial	
www.le.ac.uk/so/css/resources/mercurycolumns/jw54.html	www.leicestermercury.co.uk/index.jsp		Educational	
www.leeds.ac.uk/cath/congress/2002/programme/access.html	www.ahrb.ac.uk/		Administrative	
www.lincoln.ac.uk/home/aiminghigher/	www.lincolnshire.gov.uk/learning/learning.htm		Professional	
www.csc.liv.ac.uk/~ctag/research.html	194.66.183.26/WEBSITE/GOW/ViewGrant.ASPx?Grant= GR/N09855/01&bannerlink=Programme%20support		Research Oriented	

cwis.livjm.ac.uk/olf/oee/publications/hudd.html	www.statcounter.com		Superficial	
www.lmu.ac.uk/health/hhs/groups.htm	www.leedsmet.ac.uk/city/index.htm		Administrative	
www.londonmet.ac.uk/services/sas/library-services /subject-help/subjects/biology/bi3.cfm	www.nature.com/nri/info/info_scope.html		Research Oriented	
www.lse.ac.uk/collections/BSPS/links.htm	www.src.uchicago.edu/prc/	Educational	Educational	
www.cs.man.ac.uk/~jls/CS6482/postwork/index.html	www-anw.cs.umass.edu/~rich/book/the-book.html		Professional	
www.lr.mdx.ac.uk/lib/subjects/computing/resources/databases.htm	www.cmpnet.com/publist	Technical	Technical	
www.library.mmu.ac.uk/eresource/lawuk.html	www.guide-on-line.lawsociety.org.uk		Research Oriented	
www.ncl.ac.uk/cpact/bl.html	www.eigenvector.com/		Technical	
ecospace.newport.ac.uk/ecology2/ecol2week.htm	validator.w3.org/check/referer		Superficial	Superfic
oldweb.northampton.ac.uk/lrs/ejournals/titles/pageJ.html	swets2.nesli.ac.uk/link/access_db?issn=1042-9573		Research Oriented	
www.ntu.ac.uk/careers/student/yoursubject/business/2225gp.html	www.iii.co.uk/	Professional	Professional	
library.open.ac.uk/aboutus/opal/techGuide.html	www.cetus-links.org/		Educational	
users.ox.ac.uk/~quarrell/new.html	www.contracostatimes.com/mld/cctimes/news/local/states/california/counties/alameda_county/cities_neighborhoods/montclair/11946548.htm		Social/leisure	
library.paisley.ac.uk/ejournals/JoD-JoN.htm	journalsonline.tandf.co.uk/link.asp?id=104289		Research Oriented	
www.port.ac.uk/departments/academic/maths/infoforstaff/	www.nag.co.uk/numeric/numerical_libraries.asp		Technical	
www.qmced.ac.uk/buscomm/buscomm_links.htm	www.dti.gov.uk/		Administrative	
www.chem.qmul.ac.uk/iubmb/enzyme/EC2/4/1/195.html	www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=89271016&dopt=Abstract		Research Oriented	
www.qub.ac.uk/cm/cb/ennis6.htm	www.bsaci.org		Professional	
www.ams.rdg.ac.uk/microbiology/ABC_transporters /S_Meliloti_ABC.htm	sequence.toulouse.inra.fr/dna.script/meth.DNA.DrawMap.cgi.ok.pl?ORGA_ CONFIG=/www/Common/melilo.config&ZOOM_ORF=SMa1753		Research Oriented	
www2.rgu.ac.uk/library/resource/journals/r.htm	search.epnet.com/direct.asp?db=buh&jid=%22ER3%22&scope=site		Research Oriented	Other
www.rhul.ac.uk/Research-and-Enterprise /EnterpriseSupportCentre/ktp.html	www.ktponline.org.uk		Professional	
www.dcs.shef.ac.uk/~ajhs/publish.html	www.jot.fm/issues/issue_2004_01/column4		Research Oriented	
www.shu.ac.uk/schools/cms/teaching/pc/JDC /rmi/exercises/SimpleBankingSystem/index.html	java.sun.com/index.html		Superficial	

mercury.soas.ac.uk/Unison/links.html	www.dtlr.gov.uk/		Professional	
www.sng.ecs.soton.ac.uk/mailscanner/serve/cache/42.html	www.eicar.org/anti_virus_test_file.htm		Technical	
www.staffs.ac.uk/courses/gateway/advice_studies /how/assess_strat/index.php	staffsunion.com/	Social/leisure	Social/leisure	
www.strath.ac.uk/Departments/specneeds/advice_blind.html	508AS.usablenet.com/508AS/1.2.1.1/help/UsableNetApproved.h tml		Other	
www.sunderland.ac.uk/~wl0lfe/framebookmarks /newcomplete.htm	www.synapse.net/~woodall/html.htm		Professional	Professi
www.sunderland.ac.uk/~ts0jti/bookmarks/java.htm	www.mercury.com/java-tutor/	Technical	Technical	
www-icprocessing.ee.surrey.ac.uk/4,0industry.htm	www.jameswatt.ac.uk/		Administrative	
www.sussex.ac.uk/Users/ssfb4/MPLD/mpld.htm	www.uwc.edu/fonddulac/faculty/rrigteri/biomed.htm		Research Oriented	
astronomy.sussex.ac.uk/~sjo/idl_help/libs.html	astro.ic.ac.uk/elais/help/text/fits_bintable.ps		Technical	
sudsnet.abertay.ac.uk/ConveyanceSwale2.htm	www.epsrc.ac.uk/website/index.aspx		Administrative	
www.tees.ac.uk/depts/lis/internet/internetmeta.cfm	www.surfwax.com		Technical	
vr.tees.ac.uk/VRCentral_feb03/Trinity.htm	www.trinitygardens.co.uk		Social/leisure	
www.uclan.ac.uk/library/usersupport/lrs/subjects /eresources/music/multicomposer.htm	www.medieval.org/emfaq/composers/	Personal	Personal	Researc Oriente
www.uclan.ac.uk/library/usersupport/lrs/subjects /eresources/spanisher.htm	www.uky.edu/Subject/latinamlit.html		Technical	
www.uea.ac.uk/~gs692/htmfiles/bottom.htm	www.ukarabicpages.co.uk		Technical	
homepages.uel.ac.uk/u0116401/switchDefinition.htm	searchnetworking.techtarget.com/sDefinition/0,,sid7_gci213079, 00.html		Navigational	
www.ulst.ac.uk/library/arts/politics/A-Z-list.htm	www.yale.edu/gsp/east_timor/index.html		Research Oriented	
www.ess.uwe.ac.uk/Kosovo/Kosovo-Ethnic%20Cleansing12c.htm	www.alb-net.com/index.htm		Professional	
www.uwic.ac.uk/uwicnet/studserv/Careers/links_alternatives.htm	www.activitycamps.com/		Social/leisure	
www.uwic.ac.uk/library/information/subjects /llandaff/biomedicalsciencesinfo.htm	www.mit.edu:8001/people/cdemello/univ.html		Administrative	
www2.warwick.ac.uk/fac/soc/law/elj/jilt/1997_3/marsden2/	www.europarl.eu.int/dg1/a4/en/a4-97/a4-0219.htm		Research Oriented	
www.scit.wlv.ac.uk/wwlib/cat.section.3.html	www.seahunt.co.uk/		Professional	
wwwedit.wmin.ac.uk/sshl/new/Politics/New_Students.htm	home/sshl/new/history/default.htm	Superficial	Superficial	
www-users.york.ac.uk/~nm15/socrates.htm	www.univ-lyon1.fr		Administrative	

Appendix 12: Number of Static Pages for New Zealand Universities 2000 - 2005

University Name	Aug-00	Feb-02	Jan-03	Dec-03	Jan-05
Auckland University	36771	67023	62295	69720	61447
Auckland University of Technology	14253	8843	7061	8575	12055
University of Canterbury	56012	59757	29128	29157	29388
Lincoln University	16766	3140	3530	3590	3839
Massey University	28463	27172	26294	25734	26591
Otago University	47345	73044	41295	43255	38564
Victoria University of Wellington	180090	36045	38478	79241	36047
Waikato University	47193	40118	39947	35594	45026
Average Number of Static Pages	53362	39393	31004	36858	31620
Median Number of Static Pages	41982	38082	33803	32376	32718

Appendix 13: Number of Static Pages for Australian Universities 2000 - 2005

University Name	Aug-00	Jan-02	Mar-03	Feb-04	Mar-05
Australian Catholic University	8873	11228	6371	4935	6193
University of Adelaide	65268	127323	104898	45686	59288
Australian National University	67255	15366	25861	141379	149986
University of Ballarat	6161	5040	5304	15775	18341
Bond University	6658	12666	12831	19283	17540
University of Canberra	26534	4956	20517	19312	21570
Edith Cowan University	19190	17669	9867	27347	29308
Central Queensland University	32928	52164	61371	31768	33098
Charles Sturt University	29718	99842	80015	66114	70744
Curtin University	11724	75177	66220	51145	69763
Deakin University	3316	43080	36321	43802	53239
Flinders University	34727	28795	26829	20960	26410
Griffith University	46149	68565	90848	72666	76352
James Cook University	632	98248	92730	65418	60450
La Trobe University	13160	65240	76089	47568	40827
Monash University	56086	177316	160416	175792	161421
Macquarie University	333	77642	50166	78066	68767
Murdoch University	1	52169	56339	37836	33541
Newcastle University	29537	74037	45907	38862	34289
Northern Territory University / CDU	27035	29037	16234	15553	14334
Queensland University of Technology	40275	82509	59110	39684	25948
Royal Melbourne Institute of Technology	38895	75666	122381	90569	56881
Southern Cross University	9797	16603	10584	14882	9374
Swinburne University of Technology	29413	24887	42732	44842	11556
University of New England, Australia	27883	23036	27558	24023	28228
University of Melbourne	56530	201916	188372	194198	159104
University of South Australia	37266	88488	29890	33583	48473

University of New South Wales	50522	251367	219317	163822	210431
University of Wollongong	30860	74210	67988	32023	50470
University of Queensland	48040	179174	94226	119158	98904
University of the Sunshine Coast	582	1280	1109	1148	1561
University of South Queensland	21366	48844	86217	82677	93550
University of Sydney	50425	200002	152870	138025	151412
University of Tasmania	46747	83734	52082	52718	50889
University of Technology, Sydney	57517	63066	43122	44775	51608
University of Western Australia	55658	121958	178139	150500	157347
University of Western Sydney	31766	35747	26428	22680	10844
Victoria University	12933	18428	26575	39365	39747
Average Number of Static Pages	29783	71749	65101	60735	60573
Median Number of Static Pages	29628	64153	51124	44289	49472

Appendix 14: Number of Static Pages for UK Universities 2000 - 2004

University Name	Jun-00	Jul-01	Jun-02	Jun-03	Jun-04
University of Aberdeen	7253	67440	103874	123566	133841
University of Wales, Aberystwyth	31886	51752	74963	92137	111930
University of Abertay, Dundee	219	1757	1131	481	1565
University of the Arts London (formerly the London Institute, created in early 2004)				1324	3796
Anglia Polytechnic University	5620	22788	14716	22915	20400
Aston University	25760	20472	12857	11734	13110
University of Wales Bangor	15472	15308	17370	22265	24095
University of Bath	8246	45267	51904	73754	64446
Bath Spa University College		1588	2014	3071	3972
Birkbeck College, University of London	26661			29879	40292
University of Birmingham		307787	260257	252166	278841
University of Bournemouth		19781	23030	29092	28439
University of Bradford	8073	65201	29697	34704	34168
University of Bristol	3977	120311	113803	121651	168002
Oxford Brookes University	2877	18008	23399	25883	28785
Brunel University	7007	70749	50706	41372	36858
University of Brighton	4603	45410	49346	43399	18332
University of Buckingham	3404	1628	1430	1459	1479
University of Cambridge		323855	206656	324795	330336
Canterbury Christ Church University College		14686	17067	8393	7067
Chichester College		239	211	355	489
Cardiff University	44524	54039	44351	47246	47269
City University, London	9109	63856	66336	46493	45647
Cranfield University	10041		9691	13324	15991
University of Coventry	10469	14541	9575	11205	11448
University of Derby	10651	17800	11841	10805	7677

De Montfort University	10968	59070	59917	1271	1042
University of Dundee	12787	33464	51331	68871	74498
University of Durham	13573	69544	77006	94775	72673
Edinburgh College of Art				226	76
University of Edinburgh	15645	182965	229192	266073	224629
University of Essex	16822	51368	52333	55126	49210
University of Exeter	7915	100519	73500	77766	88130
Goldsmiths College, University of London	1288	9692	9989	12697	8925
Glasgow Caledonian University		9392	5551	8232	6887
University of Glasgow		166392	367140	136197	149117
University of Glamorgan		11487	14626	10379	8545
University of Gloucestershire			1488	1509	5673
University of Greenwich	7605	21516	21676	23994	17769
Glasgow School of Art				2406	27
Homerton College Cambridge				227	522
Harper-Adams Agricultural College		385	520	911	831
University of Hertfordshire	12612	36642	63321	21599	20836
University of Huddersfield	13425	18352	10918	28834	10898
University of Hull		17281	15619	16909	18304
Heriot-Watt University	6444	56491	60396	55307	55126
Imperial College, University of London		262906	177225	312711	107762
Institute of Cancer Research, University of London	30287			1190	1457
Institute of Education, University of London	3065			2472	2413
King’s College London	32971	76801	59037	44471	52884
University of Keele	19101	23996	16809	18601	20559
Kingston University	3075	18751	13594	6384	6275
University of Wales, Lampeter	36751	3424	3497	3997	2961
University of Lancaster		199209	71544	79894	80251
University of Lincoln and Humberside	824	8552	1405	13952	13491

University of Loughborough	8499	46710	34203	48274	38682
University of Leicester	13554	72840	34492	40386	49352
University of Leeds	17033	135231	147827	118518	107340
London University Business School				4657	2435
London Metropolitan University (created in August 2002 by merging the University of North London and London Guildhall University)	11735	30001	62590	36979	40276
University of Liverpool	6722	26634	18262	20355	29672
Liverpool John Moores University	13879	21539	30525	37614	14545
Leeds Metropolitan University	6973	11090	10490	8213	10507
London School of Economics	21703	24813	29757	36929	44067
London School of Hygiene and Tropical Medicine	1344			3920	4053
University of Luton	1461	4085	1338	1422	1742
University of Manchester		69461	124093	130099	107806
University of Middlesex		50047	41933	36057	31710
Manchester Metropolitan University		36025	21108	27604	25977
Napier University	10179	89817	98712	46387	33921
University of Newcastle	313	122764	116941	112439	140036
University of Wales College Newport	903	2625	2949	3731	4202
Northampton University College		4750	6666	8435	6575
University of Nottingham	10245	100913	70042	74639	87679
Nottingham Trent University		19719	19491	19719	18909
Open University	10182	51943	45584	43439	48167
University of Oxford		225384	197204	227224	275784
University of Paisley	3108	27268	122	136	2611
University of Plymouth	5319	21148	23703	16338	11440
University of Portsmouth	7709	53212	70267	85060	91555
Queen Margaret University College		7453	8804	10746	19165
Queen Mary, University of London	8563	41498	23340	22209	14580
Queen’s University Belfast	12437	71076	54601	51818	54120

The Royal College of Art				3940	6860
University of Reading	10900	89765	91278	94207	74695
The Robert Gordon University	8456	8405	7589	5567	5276
Royal Holloway, University of London	26862	37311	42738	23801	25725
University of Surrey, Roehampton				5267	9272
The Royal Vetinary College, University of London				768	945
University of Salford	18818	15582	8246	18232	13998
South Bank University	16712	35295	19067	20896	
St.George's Hospital Medical School				4881	2787
University of Sheffield	2797	97261	44142	101744	108667
Sheffield Hallam University	12734	45925	56375	53085	31486
School of Oriental and African Studies	1866	2974	5681	5802	4287
University of Southampton	55912	111568	121726	199815	136941
University of St Andrews	13237	140602	83766	36157	83017
University of Staffordshire		63439	35756	42424	39828
University of Stirling	17749	24026	24312	24503	27101
University of Strathclyde	15539	95724	80482	59485	54644
University of Sunderland	18973	64874	31277	22308	25509
The Surrey Institute of Art and Design University College		165	343	626	1056
University of Surrey	9611	124011	89091	96455	97550
University of Sussex	10169	78479	68030	68556	61197
University of Wales Swansea	20724	31811	27521	23190	30525
University of Teeside	2038	15954	11174	14537	9592
Thames Valley University	7001	13313	7214	11318	14026
University of Central England	3508	16186	21318	12693	25541
University College London	37541	183442	142018	199796	230330
University of Central Lancashire	11356	27741	12561	13793	16398
University of East Anglia	14557	37123	46966	60261	57255
University of East London	7332	16997	17657	20951	23392

University of Kent at Canterbury		45750	47112	32351	39822
University of London School of Pharmacy	92			104	1386
University of Ulster	34040	97950	75727	95960	61870
University of Manchester Institute of Science and Technology	8631	55528	32922	55477	44252
University of Northumbria	8664	9615	13133	38007	22033
University of the West of England	17355	19198	24725	26183	29688
University of Wales Institute at Cardiff	2341	3317	3971	4336	5336
University of Warwick		49241	47159	61017	77267
University of Wolverhampton	100629	57350	53346	34974	20039
University of Westminster	23848	18213	21174	30699	22320
University College Worcester		5347	7184	9088	29343
University of York	46955	110644	67257	88844	103317
Number of Universities	89	108	110	124	123
Average	13931	55747	49445	46419	45297
Median	10182	36334	30141	24249	25509

Appendix 15: Number of Dynamic Pages for New Zealand Universities 2001 - 2005

University Name	June 2001-May 2002	June 2002-May 2003	June 2003-May 2004	June 2004-May 2005
Auckland University	548	1425	2290	4050
Auckland University of Technology	312	742	902	5462
University of Canterbury	26	25	179	5783
Lincoln University	112	83	330	2244
Massey University	1264	723	52	44
Otago University	132	72	106	1416
Victoria University of Wellington	369	354	2329	6464
Waikato University	1744	1549	5579	11889
Average Number of Dynamic Pages	563	622	1471	4669

Appendix 16: Number of Dynamic Pages for Australian Universities 2001 - 2005

University Name	June 2001-May 2002	June 2002-May 2003	June 2003-May 2004	June 2004-May 2005
Australian Catholic University	28	54	3890	1889
University of Adelaide	177	214	683	20479
Australian National University	4189	4473	4773	5430
University of Ballarat	813	1195	1134	2557
Bond University	0	185	349	1668
University of Canberra	73	8413	3536	4862
Charles Darwin University (formally Northern Territory University)	4719	984	5777	8489
Central Queensland University	32	32	128	180
Charles Stuart University	6623	6198	13431	13402
Curtin University	4452	1325	553	490
Deakin University	3488	5466	4745	4110
Edith Cowan University	191	266	26	2617
Flinders University	1504	2363	2603	3039
Griffith University	1074	446	284	261
James Cook University	402	1059	1305	2107
La Trobe University	1914	2081	1837	1420
Monash University	502	212	152	2519
Macquarie University	56	44	44	121
Murdoch University	868	1140	5650	3066
Newcastle University	531	175	80	107
Queensland University of Technology	3068	83	301	283
Royal Melbourne Institute of Technology	1641	11738	23180	48736
Southern Cross University	644	2322	3036	3072
Swinburne University of Technology	577	450	925	597
University of New England	3	6	3	12
University of Melbourne	530	415	475	412
University of South Australia	872	883	2429	5323

University of New South Wales	55	38	38	15
University of Wollongong	288	684	947	984
University of Queensland	6814	8555	8687	13060
University of the Sunshine Coast	7	2	1	1302
University of South Queensland	765	956	900	816
University of Sydney	2263	2128	694	929
University of Tasmania	0	93	92	1380
University of Technology, Sydney	64	593	626	727
University of Western Australia	5	19608	3178	4791
University of Western Sydney	19428	1435	1937	3120
Victoria University	399	4504	4539	5231
Average Number of Dynamic Pages	1817	2390	2710	4463

Appendix 17: Number of Dynamic Pages for UK Universities 2001 - 2005

University Name	June 2001-May 2002	June 2002-May 2003	June 2003-May 2004	June 2004-May 2005
University of Aberdeen	4407	10039	12434	18935
University of Wales, Aberystwyth	1065	2230	3119	3818
Anglia Polytechnic University	133	277	198	1498
University of the Arts London (formerly the London Institute, created in early 2004)				229
Aston University	209	314	192	256
University of Wales Bangor	252	1486	2762	6332
University of Bath	1742	2832	6088	3749
Bath Spa University College	229	784	551	495
Birkbeck College, University of London			111	171
University of Birmingham	630	1717	1880	1005
University of Bournemouth	6	42	204	148
University of Bradford	1731	1348	1744	2179
University of Bristol	2535	2378	3925	16540
Oxford Brookes University	135	383	481	1708
Brunel University	6279	4398	4216	4207
University of Brighton	5	5	0	0
University of Buckingham	0	0	0	8
University of Cambridge	1092	844	1150	1386
Canterbury Christ Church University College	0	0	5	40
Cardiff University	1727	472	717	1428
Chichester College	750	67	110	124
City University, London	794	123	189	228
University of Coventry	598	20	32	1700
Cranfield University		34	113	271
University of Derby	106	742	984	3623
De Montfort University	13	867	1259	786

University of Dundee	127	156	200	319
University of Durham	6106	11609	17639	16088
Edinburgh College of Art			72	57
University of Edinburgh	132	146	400	1304
University of Essex	4	15	145	537
University of Exeter	2454	1753	2550	3828
Glasgow Caledonian University	624	1	2	1
University of Glasgow	1744	1010	869	727
University of Glamorgan	220	179	1444	1608
University of Gloucestershire		38	97	44541
Goldsmiths College, University of London	0	87	600	601
University of Greenwich	2166	1975	2243	122
Glasgow School of Art			38	3
Harper-Adams Agricultural College	0	222	300	403
Homerton College Cambridge			0	68
University of Hertfordshire	1717	1108	820	156
University of Huddersfield	432	150	77	284
University of Hull	42	20	91	282
Heriot-Watt University	15	35	38	2
Imperial College, University of London	1110	696	5075	991
Institute of Cancer Research, University of London			40	24
Institute of Education, University of London			1	1
King’s College London	5089	6367	792	1358
University of Keele	6	0	5	49
University of Kent at Canterbury	1557	3612	53	6119
Kingston University	882	657	480	0
University of Wales, Lampeter	0	4	4	0
University of Lancaster	4	85	113	158
University of Loughborough	907	546	312	686

University of Leicester	7576	6440	6526	6390
University of Leeds	1732	1778	2915	3264
University of Lincoln and Humberside	5	0	550	736
University of Liverpool	2	52	21	7
Liverpool John Moores University	283	370	195	9584
Leeds Metropolitan University	10150	46	29	12
London University Business School			9	9
London Metropolitan University (created in August 2002 by merging the University of North London and London Guildhall University)	17536	20410	11827	2153
South Bank University	512	750	1269	643
London School of Economics	1747	17953	27146	46767
London School of Hygiene and Tropical Medicine			1875	2103
University of Luton	5182	550	608	822
University of Manchester	65	70	60	87
University of Middlesex	0	0	0	0
Manchester Metropolitan University	17	218	856	4465
Napier University	499	14108	1203	1871
University of Newcastle	1505	2448	14914	39505
University of Wales College Newport	0	0	0	0
Northampton University College	10	442	947	155
University of Northumbria	416	0	0	10
University of Nottingham	374	858	1057	509
Nottingham Trent University	49	312	732	3457
Open University	65	69	420	598
University of Oxford	424	351	324	203
University of Paisley	49	4	4	675
University of Plymouth	170	829	642	159
University of Portsmouth	0	0	0	20
Queen Margaret University College	3	93	164	167

Queen Mary, University of London	0	8	20	0
Queen's University Belfast	4589	1204	1189	1335
The Royal College of Art			0	0
University of Reading	1263	2575	1591	298
The Robert Gordon University	2	336	346	653
Royal Holloway, University of London	82	127	48	1480
University of Surrey, Roehampton	4	177	8835	9681
The Royal Vetinary College, University of London			18	0
University of Salford			534	836
St.George's Hospital Medical School			140	39
University of Sheffield	4523	771	4451	7186
Sheffield Hallam University	1888	2361	3419	4475
School of Oriental and African Studies	0	92	670	363
University of Southampton	1408	1152	1106	1702
University of Staffordshire	541	270	1628	1118
University of St Andrews	14	464	221	375
University of Stirling	303	221	217	308
University of Strathclyde	14775	3047	269	198
University of Sunderland	8	11	20	12
The Surrey Institute of Art and Design University College	164	2	68	69
University of Surrey	530	2862	1328	1134
University of Sussex	505	506	945	1
University of Wales Swansea	371	358	15	78
University of Abertay, Dundee	91	122	283	135
University of Teeside	1054	1833	1995	2596
Thames Valley University	0	0	1	170
University of Central England	24	14	1519	557
University College London	209	572	2886	3260
University of Central Lancashire	2137	201	120	1086

University of East Anglia	148	118	146	134
University of East London	134	79	159	18
University of London School of Pharmacy			0	0
University of Ulster	51	337	1640	3454
University of Manchester Institute of Science and Technology	55	600	753	2039
University of the West of England	3013	1812	2567	4144
University of Wales Institute at Cardiff	308	360	178	447
University of Warwick	855	1157	1058	798
University of Wolverhampton	233	88	20	76
University of Westminster	248	661	940	6697
University College Worcester	11	15	6	48376
University of York	192	282	172	268
Average Number of Dynamic Pages	1277	1407	1567	3074

Appendix 18: Number of Non-HTML Pages for New Zealand Universities 2001 – 2005

University Name	June 2001-May 2002	June 2002-May 2003	June 2003-May 2004	June 2004-May 2005
Auckland University	15051	17557	25911	20480
Auckland University of Technology	1535	1823	2786	3759
University of Canterbury	10950	9038	10600	11204
Lincoln University	251	377	512	607
Massey University	5048	5791	7136	9147
Otago University	22675	13500	17687	11982
Victoria University of Wellington	10937	13581	33216	10095
Waikato University	8125	11759	14559	15808
Average Number of Non-HTML Pages	9322	9178	14051	10385

Appendix 19: Number of Non-HTML Pages for Australian Universities 2001 – 2005

University name	June 2001-May 2002	June 2002-May 2003	June 2003-May 2004	June 2004-May 2005
Australian Catholic University	2438	1181	1342	1568
University of Adelaide	42135	41872	15800	23216
Australian National University	2301	5447	58344	70899
University of Ballarat	1384	2215	7268	9227
Bond University	2953	3609	4569	3745
University of Canberra	359	6067	6083	3690
Edith Cowan University	8121	5248	10131	4040
Central Queensland University	9554	15759	9384	12456
Charles Stuart University	18771	13510	14936	11647
Curtin University	14310	17074	18144	19352
Deakin University	7215	3149	5698	32303
Flinders University	6590	6966	4677	8719
Griffith University	20864	32543	40743	6042
James Cook University	27295	28682	26780	46533
La Trobe University	20927	25208	16141	22972
Monash University	57260	55565	68451	16573
Macquarie University	21284	18240	33782	64289
Murdoch University	12927	12385	9987	32752
Newcastle University	18398	10863	10298	10770
Charles Darwin University (formally Northern Territory University)	7202	4896	4750	10947
Queensland University of Technology	24725	14677	12119	8107
Royal Melbourne Institute of Technology	14844	16311	45514	11657
Southern Cross University	2103	1622	3321	2067

Swinburne University of Technology	4795	12130	18053	12355
University of New England	4126	6392	5846	8117
University of Melbourne	41892	51407	84769	61184
University of South Australia	24191	7424	12630	22405
University of New South Wales	78509	91318	79568	109452
University of Wollongong	25247	20358	8973	20006
University of Queensland	48902	28583	43339	35861
University of the Sunshine Coast	139	145	166	155
University of South Queensland	6987	21070	29540	33066
University of Sydney	67211	41008	54232	62354
University of Tasmania	17925	14575	18666	21077
University of Technology, Sydney	13251	7736	11475	16150
University of Western Australia	37912	35293	46800	52055
University of Western Sydney	9555	7315	6261	2714
Victoria University	4970	7571	12207	12965
Average Number of Non-HTML Pages	19199	18300	22652	23776

Appendix 20: Number of Non-HTML Pages for UK Universities 2001 – 2005

University Name	June 2001-May 2002	June 2002-May 2003	June 2003-May 2004	June 2004-May 2005
University of Aberdeen	12571	29740	23249	53618
University of Wales, Aberystwyth	6698	15303	41959	32752
University of Abertay, Dundee	83	59	225	250
University of the Arts London (formerly the London Institute, created in early 2004)			132	340
Anglia Polytechnic University	5407	2722	6193	7016
Aston University	4398	2562	2422	3489
University of Wales Bangor	2078	3058	5740	6036
University of Bath	9837	14354	21268	26231
Bath Spa University College	145	213	642	1251
Birkbeck College, University of London			6752	8381
University of Birmingham	35573	33733	42104	184878
University of Bournemouth	1397	1858	3502	4489
University of Bradford	11094	6579	9956	9747
University of Bristol	29391	33938	38078	82674
Oxford Brookes University	4068	6267	7543	6002
Brunel University	14126	12757	9690	12327
University of Brighton	3048	3139	3128	5857
University of Buckingham	98	31	34	131
University of Cambridge	103850	65227	123209	122663
Canterbury Christ Church University College	1992	3766	1136	1721
Chichester College	47	43	172	296
Cardiff University	12372	9296	11136	12257
City University, London	13543	13220	10419	12112
Cranfield University		1017	1881	2578
University of Coventry	2197	1915	1994	2470
University of Derby	3501	2122	1815	1820

De Montfort University	25170	18998	170	98
University of Dundee	5703	7450	8090	12906
University of Durham	28264	27869	45837	32274
Edinburgh College of Art			2	27
University of Edinburgh	42967	56157	89957	82164
University of Essex	12876	13485	11884	19112
University of Exeter	30854	22840	26333	38701
Goldsmiths College, University of London	1095	1618	3188	3543
Glasgow Caledonian University	865	710	836	1838
University of Glasgow	53115	255935	40047	60284
University of Glamorgan	2322	3010	2447	2837
University of Gloucestershire		111	124	496
University of Greenwich	5471	5916	6395	4985
Glasgow School of Art			82	0
Homerton College Cambridge			151	235
Harper-Adams Agricultural College	5	113	207	183
University of Hertfordshire	4377	11114	4159	5690
University of Huddersfield	5683	1920	9288	2103
University of Hull	2833	2588	2911	3921
Heriot-Watt University	18261	20347	18685	20129
Imperial College, University of London	74050	75531	147001	54037
Institute of Cancer Research, University of London			259	329
Institute of Education, University of London			474	445
King’s College London	25992	19881	13210	17186
University of Keele	2468	2644	4104	5308
Kingston University	2243	2484	1685	1853
University of Wales, Lampeter	334	506	807	547
University of Lancaster	80675	17527	21850	24644
University of Lincoln and Humberside	2353	184	8826	4971

University of Loughborough	7569	6938	5274	10519
University of Leicester	30970	9108	12388	18575
University of Leeds	33185	33286	19446	27558
London University Business School			1315	1065
London Metropolitan University (created in August 2002 by merging the University of North London and London Guildhall University)	1841	4109	4113	5584
University of Liverpool	12239	8566	10251	19828
Liverpool John Moores University	4206	11165	17857	4718
Leeds Metropolitan University	743	919	1256	1505
London School of Economics	10555	11016	15154	22010
London School of Hygiene and Tropical Medicine			870	868
University of Luton	1018	195	156	336
University of Manchester	12627	17050	50272	44932
University of Middlesex	6591	9060	8108	9286
Manchester Metropolitan University	11040	5506	7510	7183
Napier University	25442	38042	13236	11473
University of Newcastle	20950	22745	24017	41572
University of Wales College Newport	590	614	694	1215
Northampton University College	426	1118	1539	969
University of Nottingham	34159	25748	30028	40490
Nottingham Trent University	2466	3721	3323	3441
Open University	10441	8082	8239	9350
University of Oxford	63615	69291	82122	112505
University of Paisley	13318	22	19	716
University of Plymouth	4055	7551	6315	5814
University of Portsmouth	9797	11586	14888	20655
Queen Margaret University College	2007	1538	2053	9338
Queen Mary, University of London	11929	6340	6730	5332
Queen’s University Belfast	25225	10748	13641	14279

The Royal College of Art			190	328
University of Reading	36952	38953	43214	34959
The Robert Gordon University	1772	1495	2157	2565
Royal Holloway, University of London	5173	7381	8169	10337
University of Surrey, Roehampton			1185	4383
The Royal Vetinary College, University of London			205	298
University of Salford	2800	1434	4548	4074
South Bank University	11429	3628	3722	
St.George's Hospital Medical School			653	661
University of Sheffield	17086	8991	26068	31692
Sheffield Hallam University	8612	14000	19396	7052
School of Oriental and African Studies	401	1130	1716	1673
University of Southampton	28139	38717	77306	45123
University of St Andrews	28663	19570	3806	23691
University of Staffordshire	23148	6861	10803	12204
University of Stirling	4188	4613	4973	10593
University of Strathclyde	25395	21787	15987	15021
University of Sunderland	20203	8467	7667	9736
The Surrey Institute of Art and Design University College	5	15	95	256
University of Surrey	19836	15036	19716	62890
University of Sussex	37265	18274	17297	22252
University of Wales Swansea	8342	8130	6114	9329
University of Teeside	3659	2745	3919	4250
Thames Valley University	731	1199	1418	1872
University of Central England	4551	7137	1615	10900
University College London	67271	38888	69144	111376
University of Central Lancashire	7783	2979	3737	4904
University of East Anglia	8034	10758	15502	18371
University of East London	3377	3886	7936	8262

University of Kent at Canterbury	8384	9111	9522	13682
University of London School of Pharmacy			31	396
University of Ulster	13553	17520	21980	17335
University of Manchester Institute of Science and Technology	12851	9345	12031	14302
University of Northumbria	1287	1803	17768	6198
University of the West of England	1760	3529	4885	8239
University of Wales Institute at Cardiff	319	763	1188	1826
University of Warwick	12704	13375	19641	25798
University of Wolverhampton	9172	7561	7735	7259
University of Westminster	1628	2647	2807	3574
University College Worcester	710	1175	1995	8180
University of York	41782	17725	25143	37911
Average Number of Non-HTML Pages	14272.76852	13877.71818	13767.62903	16805.69106

Appendix 21: Number of Inlinks for New Zealand Universities 2000 - 2005

University Name	Jul-00	Feb-02	Jan-03	Dec-03	Jan-05
Auckland University	0	684	754	1006	850
Auckland University of Technology	5	52	56	87	83
University of Canterbury	125	368	415	364	380
Lincoln University	51	266	281	191	183
Massey University	111	377	370	402	365
Otago University	58	277	322	397	380
Victoria University of Wellington	137	471	583	638	636
Waikato University	192	479	477	557	476
Average Number of Inlinks	85	372	407	455	419

Appendix 22: Number of Inlinks for Australian Universities 2000 - 2005

University Name	Aug-00	Jan-02	Mar-03	Feb-04	Mar-05
Australian Catholic University	140	304	252	312	306
University of Adelaide	2082	2519	2364	2261	2590
Australian National University	3932	8861	7257	6632	6399
University of Ballarat	152	305	270	308	321
Bond University	179	373	313	298	310
University of Canberra	743	1301	1036	1163	980
Charles Darwin University (formally Northern Territory University)	273	496	525	576	532
Edith Cowan University	642	2463	1106	1505	1634
Central Queensland University	385	732	578	658	1407
Charles Stuart University	1574	3414	2691	2590	2264
Curtin University	807	2317	1849	1828	2184
Deakin University	688	1543	1293	1203	1236
Flinders University	833	1666	1731	1643	1521
Griffith University	955	2057	1670	2020	1424
James Cook University	570	1114	900	869	919
La Trobe University	869	1838	1563	1540	1571
Monash University	2081	4825	4106	3700	3740
Macquarie University	1161	2436	2273	2202	2115
Murdoch University	1078	2936	2297	2567	2334
Newcastle University	866	1712	1931	1238	1376
Queensland University of Technology	980	2616	1527	1640	1753
Royal Melbourne Institute of Technology	780	1976	1739	1647	1944
Southern Cross University	520	1112	811	830	809
Swinburne University of Technology	610	1294	934	878	565

University of New England	519	1167	947	995	1000
University of Melbourne	6909	27527	23446	9333	9458
University of South Australia	648	1657	1447	1487	1480
University of New South Wales	1502	4126	3983	4444	4639
University of Wollongong	594	1356	1143	1121	1441
University of Queensland	1533	3590	3418	3930	4270
University of the Sunshine Coast	21	79	74	93	91
University of South Queensland	352	805	636	648	629
University of Sydney	1920	5924	4863	3936	3899
University of Tasmania	809	1634	1320	1280	1300
University of Technology, Sydney	921	3164	2928	3056	3290
University of Western Australia	964	2705	2318	2709	2752
University of Western Sydney	663	1102	864	891	917
Victoria University	83	3022	5836	434	486
Average Number of Inlinks	1062	2844	2480	1960	2051

Appendix 23: Number of (Page ADM) Inlinks for UK Universities 2000 - 2005

University Name	Jul-00	Jul-01	Jul-02	Jun-03	Jun-04	Jul-05
University of Aberdeen	1153	2496	2322	2276	2207	2235
University of Wales, Aberystwyth	1152	2122	1940	1838	1698	1417
University of Abertay, Dundee		257	246	235	225	216
University of the Arts London (formerly the London Institute, created in early 2004)				273	284	278
Anglia Polytechnic University	639	1385	1221	1204	1049	566
Aston University	418	1033	834	991	858	779
University of Wales Bangor	702	1434	1369	1383	1471	1317
University of Bath	1665	3618	3009	2957	2572	2597
Bath Spa University College		145	176	155	136	157
Birkbeck College, University of London	1074			2011	1875	1750
University of Birmingham		11476	12539	7128	6377	5978
University of Bournemouth		742	674	693	698	661
University of Bradford	638	1445	1365	1270	1090	997
University of Bristol	2823	6775	5881	5973	5617	5708
Oxford Brookes University	501	1172	1121	1298	1337	1231
Brunel University	579	1311	1241	1232	1100	1340
University of Brighton	672	1380	1275	1260	1133	1063
University of Buckingham	18	36	41	47	41	45
University of Cambridge		38225	16875	45431	46132	44146
Canterbury Christ Church University College		231	242	250	272	257
Chichester College		16	15	15	15	14
Cardiff University	275	3563	3350	3268	2714	2674
City University, London	3906	1841	3041	2996	1947	2162
Cranfield University	532		1548	1278	1106	
University of Coventry	237	525	537	570	1418	553
University of Derby	329	663	587	502	425	383
De Montfort University	940	2110	2094	1971	1906	1785
University of Dundee	891	2076	1757	1777	1759	1763

University of Durham	1443	4166	3804	3963	3861	4054
Edinburgh College of Art				110	85	
University of Edinburgh	6610	19509	15291	14878	12927	11076
University of Essex	1561	3329	3026	2966	2830	2855
University of Exeter	1501	3365	3542	3005	2872	2657
Goldsmiths College, University of London	444	771	664	1433	578	606
Glasgow Caledonian University		731	750	758	793	859
University of Glasgow		9254	8316	8399	8052	7819
University of Glamorgan		412	456	490	483	452
University of Gloucestershire			285	309	326	299
University of Greenwich	269	597	458	452	402	474
Glasgow School of Art				168	148	
Homerton College Cambridge				1848	1967	70
Harper-Adams Agricultural College		51	53	50	52	43
University of Hertfordshire	643	1168	1112	1143	1072	1071
University of Huddersfield	384	795	926	712	901	1085
University of Hull		2111	2220	1886	1705	1993
Heriot-Watt University	2498	5856	4940	4860	4299	3988
Imperial College, University of London		23991	21430	21412	7629	7231
Institute of Cancer Research, University of London	93			215	208	
Institute of Education, University of London	244			657	641	
King’s College London	1502	3136	3376	3651	4458	4118
University of Keele	820	1646	1516	1451	1266	1132
Kingston University	271	586	585	573	549	477
University of Wales, Lampeter	271	424	309	300	287	266
University of Lancaster		3556	3333	3368	3238	3279
University of Lincoln and Humberside	49	137	233	216	189	166
University of Loughborough	871	1877	1814	2124	2310	2187
University of Leicester	1465	3837	4051	4154	3919	3654
University of Leeds	4180	12719	10644	11553	10687	10607

London University Business School				87	122	
London Metropolitan University (created in August 2002 by merging the University of North London and London Guildhall University)	838	189569	337790	209380	1406	1821
University of Liverpool	1815	3916	3694	4073	3711	3509
Liverpool John Moores University	454	900	857	866	840	686
Leeds Metropolitan University	222	512	503	457	473	411
London School of Economics	766	1939	1761	1912	2000	2082
London School of Hygiene and Tropical Medicine	283			442	416	
University of Luton	138	230	235	236	175	128
University of Manchester		10525	8881	10094	11174	8404
University of Middlesex		774	745	678	720	676
Manchester Metropolitan University		1583	1608	1653	1513	1201
Napier University	559	1694	1617	1423	1237	1086
University of Newcastle	2505	5402	4844	4841	4444	3985
University of Wales College Newport	128	125	124	105	304	265
Northampton University College		74	90	115	95	99
University of Nottingham	1948	8810	6743	9227	10376	10411
Nottingham Trent University		1275	1235	1296	1216	1100
Open University	1418	3526	3630	3645	3364	3122
University of Oxford		23855	15963	16541	21388	21112
University of Paisley	170	397	474	468	433	421
University of Plymouth	496	1148	1030	1101	1036	1014
University of Portsmouth	453	945	889	900	827	723
Queen Margaret University College		421	487	345	356	241
Queen Mary, University of London	1197	3443	2812	2557	2606	2917
Queen’s University Belfast	1077	2133	2216	1950	1788	1657
The Royal College of Art				146	156	
University of Reading	1000	9071	4281	4015	3721	3456
The Robert Gordon University	350	863	756	941	755	673
Royal Holloway, University of London	515	1311	1410	1644	1638	

University of Surrey, Roehampton				253	251	
The Royal Vetinary College, University of London				306	309	1713
University of Salford	525	1394	1310	1674	1277	800
South Bank University	634	1640	1382	1310		1007
St.George's Hospital Medical School				293	306	
University of Sheffield	3357	7305	6386	6451	5986	5523
Sheffield Hallam University	656	1559	1355	1526	1512	1272
School of Oriental and African Studies	251	391	428	516	562	569
University of Southampton	5633	14633	18176	19436	14809	12193
University of St Andrews	1125	4512	4277	4851	4618	4677
University of Staffordshire		896	837	816	787	783
University of Stirling	833	1746	1525	1499	1472	1465
University of Strathclyde	1509	4247	3726	3303	3224	3386
University of Sunderland	491	1061	1041	922	820	682
The Surrey Institute of Art and Design University College		76	62	50	49	44
University of Surrey	1413	3335	3060	3062	2838	2523
University of Sussex	1266	3466	3044	3224	3109	2945
University of Wales Swansea	1013	1735	1580	1876	1604	1483
University of Teeside	201	543	643	505	589	428
Thames Valley University	81	146	133	147	126	124
University of Central England	180	344	284	360	340	341
University College London	3892	48922	13395	12981	11495	11082
University of Central Lancashire	242	509	454	472	436	428
University of East Anglia	1292	2701	2371	2493	2242	2172
University of East London	317	565	464	422	347	318
University of Kent at Canterbury	1229	2707	2527	2579	2552	2560
University of London School of Pharmacy	12			36	36	
University of Ulster	828	1792	1590	1703	1706	1507
University of Manchester Institute of Science and Technology	674	2370	2290	2598	3321	

University of Northumbria	640	1727	1392	1210	1192	1086
University of the West of England	365	942	909	1043	1003	1085
University of Wales Institute at Cardiff	110	168	190	204	186	191
University of Warwick		6178	5122	5406	5133	4884
University of Wolverhampton	1460	2733	2322	2225	2166	1636
University of Westminster	502	1120	763	690	641	585
University College Worcester		124	130	239	176	143
University of York	1999	5721	5311	5460	5267	5392
Number of Universities	89	108	110	124	123	111
Average Number of Inlinks	1049	5480	5924	4521	2658	2710
Average Number of Inlinks without lmu.ac.uk	1039	3725	2853	2833	2647	2694

Appendix 24: Number of (Directory ADM) Inlinks for UK Universities 2000 - 2005

University Name	Jul-00	Jul-01	Jul-02	Jun-03	Jun-04	Jul-05
University of Aberdeen	1056	2293	2063	1964	1831	1742
University of Wales, Aberystwyth	859	1629	1504	1406	1263	1128
University of Abertay, Dundee	261	209	205	193	187	180
Anglia Polytechnic University	409	804	688	703	600	450
Aston University	356	769	635	709	624	553
University of Wales Bangor	561	1084	994	1013	964	831
University of Bath	1282	2612	2206	2145	1940	1887
Bath Spa University College		126	141	139	120	125
University of Birmingham		7982	8568	5552	4990	4712
University of Bournemouth		568	494	527	501	502
University of Bradford	567	1200	980	999	866	802
University of Bristol	2366	5394	4654	4703	4450	4325
Oxford Brookes University	401	932	907	1054	1051	991
Brunel University	692	1495	1317	1255	1100	1046
University of Brighton	505	1104	987	977	879	810
University of Buckingham	16	37	42	48	39	41
University of Cambridge		13573	11561	11979	11352	11042
Canterbury Christ Church University College		211	204	211	210	193
Chichester College		16	13	13	13	12
Cardiff University	1586	2859	2649	2543	2120	2005
City University, London	668	1478	1490	1470	1309	1151
Cranfield University	465		1312	1021	828	
University of Coventry	203	432	435	472	445	431
University of Derby	289	561	527	451	361	328
De Montfort University	750	1482	1386	1296	1175	1061
University of Dundee	870	1795	1597	1644	1536	1433
University of Durham	1221	3354	2908	2990	2941	2873
University of Edinburgh	4967	11636	10191	9827	8686	8117
University of Essex	1275	2653	2406	2253	2103	2050
University of Exeter	1252	2783	2957	2465	2238	2084
Goldsmiths College, University of London	354	597	501	569	464	496

Glasgow Caledonian University		514	540	555	497	511
University of Glasgow		6406	5898	5867	5533	5270
University of Glamorgan		367	310	334	316	284
University of Gloucestershire			240	273	281	258
University of Greenwich	231	481	412	403	372	362
Harper-Adams Agricultural College		245	363	689	41	39
University of Hertfordshire	521	972	878	884	798	777
University of Huddersfield	280	613	558	530	582	597
University of Hull		1564	1386	1407	1196	1624
Heriot-Watt University	2040	4369	3779	3659	3258	2980
Imperial College, University of London		7913	6479	6359	5696	5242
King’s College London	1104	2089	2024	2174	2160	2159
University of Keele	615	1278	1187	1169	1068	925
Kingston University	236	502	462	451	419	350
University of Wales, Lampeter	207	334	258	256	235	210
University of Lancaster		2890	2711	2678	2550	2518
University of Lincoln and Humberside	37	103	198	174	155	140
University of Loughborough	742	1571	1527	1730	1707	1634
University of Leicester	1163	2923	2767	2830	2637	2535
University of Leeds	3447	10507	8725	9321	8408	8171
University of Liverpool	1378	3032	2802	2947	2675	2511
Liverpool John Moores University	375	708	620	558	545	500
Leeds Metropolitan University	186	393	377	387	375	327
London School of Economics	532	1234	1315	1487	1522	1592
University of Luton	124	204	182	176	154	120
University of Manchester		7441	6425	6702	6819	6239
University of Middlesex		624	603	546	563	545
Manchester Metropolitan University		1234	1136	1171	1042	938
Napier University	470	1170	1083	965	921	811
University of Newcastle	2100	4356	3759	3688	3374	3120
University of Wales College Newport	90	93	91	79	101	101
Northampton University College		59	80	101	80	84
University of Nottingham	1656	4133	3708	3850	3911	3825

Nottingham Trent University		1020	1014	1022	948	849
Open University	1216	2747	2747	2798	2579	2411
University of Oxford		15687	11403	11676	10663	10102
University of Paisley	141	331	351	336	304	290
University of Plymouth	436	959	866	898	832	778
University of Portsmouth	409	818	751	784	726	630
Queen Margaret University College		254	257	245	248	195
Queen Mary, University of London	989	2614	2199	1971	2096	1966
Queen’s University Belfast	875	1733	1520	1587	1465	1352
University of Reading	844	4093	2978	2832	2680	2460
The Robert Gordon University	296	670	594	595	514	501
Royal Holloway, University of London	439	1032	1051	1129	1099	
University of Salford	445	914	795	850	764	653
South Bank University	505	1131	1002	949	846	714
University of Sheffield	2620	5777	5124	5067	4764	4439
Sheffield Hallam University	525	1197	1055	1177	1030	924
School of Oriental and African Studies	183	335	361	422	405	399
University of Southampton	3518	7887	7439	7879	7367	6486
University of St Andrews	14882	142203	72818	56200	2771	2691
University of Staffordshire		726	682	674	639	618
University of Stirling	654	1400	1253	1232	1195	1133
University of Strathclyde	1228	3137	2665	2395	2319	2263
University of Sunderland	404	833	741	711	620	506
The Surrey Institute of Art and Design University College		63	51	45	43	37
University of Surrey	1173	2636	2418	2388	2210	1991
University of Sussex	1032	2667	2406	2536	2439	2259
University of Wales Swansea	769	1329	1205	1351	1223	1126
University of Teeside	184	444	393	372	372	302
Thames Valley University	63	128	121	133	111	104
University of Central England	149	292	249	272	268	264
University College London	3029	8933	7961	8701	6661	6665
University of Central Lancashire	206	428	372	383	347	335

University of East Anglia	979	2223	1961	1961	1689	1617
University of East London	278	484	385	357	295	267
University of Kent at Canterbury	1017	2140	1992	1960	1782	1822
University of Ulster	715	1353	1254	1196	1084	913
University of Manchester Institute of Science and Technology	583	1830	1673	1859	1859	
University of Northumbria	481	1036	2982	4532	887	802
University of the West of England	296	768	727	801	762	752
University of Wales Institute at Cardiff	72	114	125	155	141	135
University of Warwick		4920	4142	4196	3867	3848
University of Wolverhampton	1218	2147	1879	1760	1566	1288
University of Westminster	408	757	620	572	522	475
University College Worcester		107	112	141	132	116
University of York	1711	4406	3886	3922	3669	3780

Appendix 25: Percentage Change in Outlinks for New Zealand Universities 2000 – 2006

University Name	Aug-00	Feb-02	Total	Non-Duplicates	Same as Previous Year	% Change	New for This Year	% Change	Lost from Previous Year	% Change
Auckland University	584	614	1198	871	327	53%	287	47%	257	42%
Auckland University of Technology	2	2441	2443	2443	0	0%	2441	100%	2	0%
University of Canterbury	0	431	431	431	0	0%	431	100%	0	0%
Lincoln University	0	50	50	50	0	0%	50	100%	0	0%
Massey University	0	170	170	170	0	0%	170	100%	0	0%
Otago University	1	432	433	433	0	0%	432	100%	1	0%
Victoria University of Wellington	0	450	450	450	0	0%	450	100%	0	0%
Waikato University	0	392	392	392	0	0%	392	100%	0	0%
		4980				7%		93%		5%

University Name	Feb-02	Jan-03	Total	Non-Duplicates	Same as Previous Year	% Change	New for This Year	% Change	Lost from Previous Year	% Change
Auckland University	614	719	1333	857	476	66%	243	34%	138	19%
Auckland University of Technology	2441	112	2553	2484	69	62%	43	38%	2372	2118%
University of Canterbury	431	420	851	542	309	74%	111	26%	122	29%
Lincoln University	50	46	96	68	28	61%	18	39%	22	48%
Massey University	170	233	403	262	141	61%	92	39%	29	12%
Otago University	432	494	926	563	363	73%	131	27%	69	14%
Victoria University of Wellington	450	471	921	570	351	75%	120	25%	99	21%
Waikato University	392	387	779	478	301	78%	86	22%	91	24%
	4980	2882				69%		31%		286%

University Name	Jan-03	Dec-03	Total	Non-Duplicates	Same as Previous Year	% Change	New for This Year	% Change	Lost from Previous Year	% Change
Auckland University	719	853	1572	994	578	68%	275	32%	141	17%
Auckland University of Technology	112	115	227	154	73	63%	42	37%	39	34%
University of Canterbury	420	387	807	523	284	73%	103	27%	136	35%
Lincoln University	46	42	88	55	33	79%	9	21%	13	31%
Massey University	233	209	442	303	139	67%	70	33%	94	45%
Otago University	494	457	951	583	368	81%	89	19%	126	28%
Victoria University of Wellington	471	459	930	641	289	63%	170	37%	182	40%
Waikato University	387	453	840	529	311	69%	142	31%	76	17%
	2882	2975				70%		30%		31%

University Name	Dec-03	Jan-05	Total	Non-Duplicates	Same as Previous Year	% Change	New for This Year	% Change	Lost from Previous Year	% Change
Auckland University	853	1002	1855	1261	594	59%	408	41%	259	26%
Auckland University of Technology	115	1263	1378	1307	71	6%	1192	94%	44	3%
University of Canterbury	387	379	766	537	229	60%	150	40%	158	42%
Lincoln University	42	41	83	57	26	63%	15	37%	16	39%
Massey University	209	311	520	361	159	51%	152	49%	50	16%
Otago University	457	507	964	606	358	71%	149	29%	99	20%
Victoria University of Wellington	459	589	1048	727	321	54%	268	46%	138	23%
Waikato University	453	381	834	541	293	77%	88	23%	160	42%
	2975	4473				55%		45%		26%

University Name	Jan-05	Jan-06	Total	Non-Duplicates	Same as Previous Year	% Change	New for This Year	% Change	Lost from Previous Year	% Change
Auckland University	1002	985	1987	1316	671	68%	314	32%	331	34%
Auckland University of Technology	1263	1298	2561	1765	796	61%	502	39%	467	36%
University of Canterbury	379	446	825	517	308	69%	138	31%	71	16%
Lincoln University	41	21	62	50	12	57%	9	43%	29	138%
Massey University	311	411	722	445	277	67%	134	33%	34	8%
Otago University	507	512	1019	609	410	80%	102	20%	97	19%
Victoria University of Wellington	589	551	1140	775	365	66%	186	34%	224	41%
Waikato University	381	346	727	442	285	82%	61	18%	96	28%
	4473	4570				69%		31%		40%

Appendix 26: Cumulative Percentage Change in Outlinks for New Zealand Universities 2000 – 2006

University Name	Feb-02	Jan-06	Total	Non-Duplicates	Same as Previous Year	% Change	New for This Year	% Change	Lost from Previous Year	% Change
Auckland University	614	985	1599	1356	243	25%	742	75%	371	38%
Auckland University of Technology	2441	1298	3739	3729	10	1%	1288	99%	2431	187%
University of Canterbury	431	446	877	755	122	27%	324	73%	309	69%
Lincoln University	50	21	71	69	2	10%	19	90%	48	229%
Massey University	170	411	581	486	95	23%	316	77%	75	18%
Otago University	432	512	944	730	214	42%	298	58%	218	43%
Victoria University of Wellington	450	551	1001	856	145	26%	406	74%	305	55%
Waikato University	392	346	738	602	136	39%	210	61%	256	74%
	4980	4570	9550	8583	967	24%	3603	76%	4013	89%

Appendix 27: Percentage Change in Outlinks for Australian Universities 2000 – 2006

University Name	Jul-00	Jan-02	Total	Non-Duplicates	Same as Previous Year	% Change	New for This Year	% Change	Lost from Previous Year	% Change
Australian Catholic University	0	708	708	708	0	0%	708	100%	0	0%
University of Adelaide	0	7406	7406	7406	0	0%	7406	100%	0	0%
Australian National University	2208	4629	6837	5296	1541	33%	3088	67%	667	14%
University of Ballarat	0	85	85	85	0	0%	85	100%	0	0%
Bond University	0	350	350	350	0	0%	350	100%	0	0%
University of Canberra	0	489	489	489	0	0%	489	100%	0	0%
Charles Darwin University (formally Northern Territory University)	671	2407	3078	2631	447	19%	1960	81%	224	9%
Central Queensland University	651	1037	1688	1143	545	53%	492	47%	106	10%
Charles Stuart University	1996	7008	9004	7790	1214	17%	5794	83%	782	11%
Curtin University	920	1815	2735	2039	696	38%	1119	62%	224	12%
Deakin University	460	1419	1879	1481	398	28%	1021	72%	62	4%
Edith Cowan University	190	1691	1881	1746	135	8%	1556	92%	55	3%
Flinders University	1371	1729	3100	2110	990	57%	739	43%	381	22%
Griffith University	761	1058	1819	1819	0	0%	1058	100%	761	72%
James Cook University	104	2122	2226	2126	100	5%	2022	95%	4	0%
La Trobe University	968	3771	4739	3952	787	21%	2984	79%	181	5%
Monash University	1983	3609	5592	4222	1370	38%	2239	62%	613	17%
Macquarie University	33	2582	2615	2589	26	1%	2556	99%	7	0%
Murdoch University	0	2571	2571	2571	0	0%	2571	100%	0	0%
Newcastle University	1200	1677	2877	1991	886	53%	791	47%	314	19%
Queensland University of Technology	2131	1416	3547	3068	479	34%	937	66%	1652	117%
Royal Melbourne Institute of Technology	1420	3701	5121	4531	590	16%	3111	84%	830	22%
Southern Cross University	1345	1492	2837	1745	1092	73%	400	27%	253	17%
Swinburne University of Technology	750	649	1399	1060	339	52%	310	48%	411	63%
University of New England	1061	1276	2337	1485	852	67%	424	33%	209	16%
University of Melbourne	749	3949	4698	4193	505	13%	3444	87%	244	6%

University of South Australia	1364	2243	3607	2638	969	43%	1274	57%	395	18%
University of New South Wales	1560	3002	4562	3492	1070	36%	1932	64%	490	16%
University of Wollongong	933	1361	2294	1699	595	44%	766	56%	338	25%
University of Queensland	1438	4565	6003	5004	999	22%	3566	78%	439	10%
University of the Sunshine Coast	0	70	70	70	0	0%	70	100%	0	0%
University of South Queensland	630	997	1627	1172	455	46%	542	54%	175	18%
University of Sydney	2979	9423	12402	10252	2150	23%	7273	77%	829	9%
University of Tasmania	1279	1640	2919	1852	1067	65%	573	35%	212	13%
University of Technology, Sydney	676	3497	4173	3994	179	5%	3318	95%	497	14%
University of Western Australia	1092	2305	3397	2645	752	33%	1553	67%	340	15%
University of Western Sydney	859	885	1744	1379	365	41%	520	59%	494	56%
Victoria University	1487	1210	2697	1869	828	68%	382	32%	659	54%
	35269	91844				28%		72%		18%

University Name	Jan-02	Mar-03	Total	Non-Duplicates	Same as Previous Year	% Change	New for This Year	% Change	Lost from Previous Year	% Change
Australian Catholic University	708	646	1354	986	368	57%	278	43%	340	53%
University of Adelaide	7406	1980	9386	7881	1505	76%	475	24%	5901	298%
Australian National University	4629	4367	8996	5705	3291	75%	1076	25%	1338	31%
University of Ballarat	85	111	196	147	49	44%	62	56%	36	32%
Bond University	350	227	577	438	139	61%	88	39%	211	93%
University of Canberra	489	971	1460	1069	391	40%	580	60%	98	10%
Charles Darwin University (formally Northern Territory University)	2407	2033	4440	2710	1730	85%	303	15%	677	33%
Central Queensland University	1037	915	1952	1290	662	72%	253	28%	375	41%
Charles Stuart University	7008	4010	11018	8260	2758	69%	1252	31%	4250	106%
Curtin University	1815	1467	3282	2252	1030	70%	437	30%	785	54%
Deakin University	1419	1373	2792	1897	895	65%	478	35%	524	38%
Edith Cowan University	1691	1815	3506	2224	1282	71%	533	29%	409	23%

Flinders University	1729	1705	3434	2119	1315	77%	390	23%	414	24%
Griffith University	1058	883	1941	1421	520	59%	363	41%	538	61%
James Cook University	2122	1887	4009	2661	1348	71%	539	29%	774	41%
La Trobe University	3771	3660	7431	4662	2769	76%	891	24%	1002	27%
Monash University	3609	3733	7342	4651	2691	72%	1042	28%	918	25%
Macquarie University	2582	2193	4775	3178	1597	73%	596	27%	985	45%
Murdoch University	2571	1696	4267	2955	1312	77%	384	23%	1259	74%
Newcastle University	1677	1120	2797	2083	714	64%	406	36%	963	86%
Queensland University of Technology	1416	1261	2677	1905	772	61%	489	39%	644	51%
Royal Melbourne Institute of Technology	3701	3546	7247	5701	1546	44%	2000	56%	2155	61%
Southern Cross University	1492	1678	3170	1843	1327	79%	351	21%	165	10%
Swinburne University of Technology	649	998	1647	1182	465	47%	533	53%	184	18%
University of New England	1276	1243	2519	1625	894	72%	349	28%	382	31%
University of Melbourne	3949	3282	7231	4876	2355	72%	927	28%	1594	49%
University of South Australia	2243	1623	3866	2684	1182	73%	441	27%	1061	65%
University of New South Wales	3002	2515	5517	3826	1691	67%	824	33%	1311	52%
University of Wollongong	1361	1982	3343	2257	1086	55%	896	45%	275	14%
University of Queensland	4565	2590	7155	5193	1962	76%	628	24%	2603	101%
University of the Sunshine Coast	70	167	237	193	44	26%	123	74%	26	16%
University of South Queensland	997	1412	2409	1635	774	55%	638	45%	223	16%
University of Sydney	9423	7419	16842	11250	5592	75%	1827	25%	3831	52%
University of Tasmania	1640	1371	3011	2119	892	65%	479	35%	748	55%
University of Technology, Sydney	3497	3697	7194	4855	2339	63%	1358	37%	1158	31%
University of Western Australia	2305	1920	4225	2882	1343	70%	577	30%	962	50%
University of Western Sydney	885	610	1495	1149	346	57%	264	43%	539	88%
Victoria University	1210	1085	2295	1416	879	81%	206	19%	331	31%
	91844	75191				66%		34%		52%

University Name	Mar-03	Feb-04	Total	Non-Duplicates	Same as Previous Year	% Change	New for This Year	% Change	Lost from Previous Year	% Change
Australian Catholic University	646	362	1008	804	204	56%	158	44%	442	122%
University of Adelaide	1980	1955	3935	2376	1559	80%	396	20%	421	22%
Australian National University	4367	4647	9014	5419	3595	77%	1052	23%	772	17%
University of Ballarat	111	166	277	205	72	43%	94	57%	39	23%
Bond University	227	284	511	334	177	62%	107	38%	50	18%
University of Canberra	971	927	1898	1179	719	78%	208	22%	252	27%
Charles Darwin University (formally Northern Territory University)	2033	16066	18099	16202	1897	12%	14169	88%	136	1%
Central Queensland University	915	977	1892	1276	616	63%	361	37%	299	31%
Charles Stuart University	4010	4637	8647	6045	2602	56%	2035	44%	1408	30%
Curtin University	1467	1348	2815	2050	765	57%	583	43%	702	52%
Deakin University	1373	1538	2911	1799	1112	72%	426	28%	261	17%
Edith Cowan University	1815	34050	35865	34750	1115	3%	32935	97%	700	2%
Flinders University	1705	1740	3445	2269	1176	68%	564	32%	529	30%
Griffith University	883	1311	2194	1654	540	41%	771	59%	343	26%
James Cook University	1887	2074	3961	2343	1618	78%	456	22%	269	13%
La Trobe University	3660	3446	7106	4138	2968	86%	478	14%	692	20%
Monash University	3733	3547	7280	4844	2436	69%	1111	31%	1297	37%
Macquarie University	2193	2577	4770	2901	1869	73%	708	27%	324	13%
Murdoch University	1696	1276	2972	2027	945	74%	331	26%	751	59%
Newcastle University	1120	1178	2298	1419	879	75%	299	25%	241	20%
Queensland University of Technology	1261	1008	2269	1742	527	52%	481	48%	734	73%
Royal Melbourne Institute of Technology	3546	3316	6862	4092	2770	84%	546	16%	776	23%
Southern Cross University	1678	1290	2968	1891	1077	83%	213	17%	601	47%
Swinburne University of Technology	998	1005	2003	1250	753	75%	252	25%	245	24%
University of New England	1243	1378	2621	1514	1107	80%	271	20%	136	10%
University of Melbourne	3282	3087	6369	4194	2175	70%	912	30%	1107	36%
University of South Australia	1623	1487	3110	1999	1111	75%	376	25%	512	34%

University of New South Wales	2515	2400	4915	3091	1824	76%	576	24%	691	29%
University of Wollongong	1982	1158	3140	2245	895	77%	263	23%	1087	94%
University of Queensland	2590	2908	5498	3711	1787	61%	1121	39%	803	28%
University of the Sunshine Coast	167	169	336	176	160	95%	9	5%	7	4%
University of South Queensland	1412	1488	2900	1730	1170	79%	318	21%	242	16%
University of Sydney	7419	6675	14094	8256	5838	87%	837	13%	1581	24%
University of Tasmania	1371	1320	2691	1743	948	72%	372	28%	423	32%
University of Technology, Sydney	3697	4536	8233	5561	2672	59%	1864	41%	1025	23%
University of Western Australia	1920	1736	3656	2473	1183	68%	553	32%	737	42%
University of Western Sydney	610	646	1256	817	439	68%	207	32%	171	26%
Victoria University	1085	1363	2448	1471	977	72%	386	28%	108	8%
	75191	121076				67%		33%		30%

University Name	Feb-04	Mar-05	Total	Non-Duplicates	Same as Previous Year	% Change	New for This Year	% Change	Lost from Previous Year	% Change
Australian Catholic University	362	436	798	549	249	57%	187	43%	113	26%
University of Adelaide	1955	1931	3886	2337	1549	80%	382	20%	406	21%
Australian National University	4647	4718	9365	5442	3923	83%	795	17%	724	15%
University of Ballarat	166	250	416	283	133	53%	117	47%	33	13%
Bond University	284	258	542	333	209	81%	49	19%	75	29%
University of Canberra	927	715	1642	1072	570	80%	145	20%	357	50%
Charles Darwin University (formally Northern Territory University)	16066	10899	26965	19483	7482	69%	3417	31%	8584	79%
Central Queensland University	977	1051	2028	1222	806	77%	245	23%	171	16%
Charles Stuart University	4637	5711	10348	6927	3421	60%	2290	40%	1216	21%
Curtin University	1348	1314	2662	1573	1089	83%	225	17%	259	20%
Deakin University	1538	1389	2927	2061	866	62%	523	38%	672	48%
Edith Cowan University	34050	1026	35076	34624	452	44%	574	56%	33598	3275%
Flinders University	1740	1887	3627	2121	1506	80%	381	20%	234	12%

Griffith University	1311	2038	3349	2500	849	42%	1189	58%	462	23%
James Cook University	2074	2392	4466	2791	1675	70%	717	30%	399	17%
La Trobe University	3446	3551	6997	4008	2989	84%	562	16%	457	13%
Monash University	3547	3212	6759	4517	2242	70%	970	30%	1305	41%
Macquarie University	2577	2522	5099	3084	2015	80%	507	20%	562	22%
Murdoch University	1276	1240	2516	1632	884	71%	356	29%	392	32%
Newcastle University	1178	1752	2930	2167	763	44%	989	56%	415	24%
Queensland University of Technology	1008	545	1553	1264	289	53%	256	47%	719	132%
Royal Melbourne Institute of Technology	3316	3286	6602	3585	3017	92%	269	8%	299	9%
Southern Cross University	1290	1181	2471	1857	614	52%	567	48%	676	57%
Swinburne University of Technology	1005	685	1690	1005	685	100%	0	0%	320	47%
University of New England	1378	1480	2858	1665	1193	81%	287	19%	185	13%
University of Melbourne	3087	2868	5955	3824	2131	74%	737	26%	956	33%
University of South Australia	1487	1616	3103	1982	1121	69%	495	31%	366	23%
University of New South Wales	2400	2692	5092	3311	1781	66%	911	34%	619	23%
University of Wollongong	1158	1360	2518	1582	936	69%	424	31%	222	16%
University of Queensland	2908	2661	5569	3620	1949	73%	712	27%	959	36%
University of the Sunshine Coast	169	186	355	233	122	66%	64	34%	47	25%
University of South Queensland	1488	1549	3037	1824	1213	78%	336	22%	275	18%
University of Sydney	6675	7240	13915	8208	5707	79%	1533	21%	968	13%
University of Tasmania	1320	1359	2679	1674	1005	74%	354	26%	315	23%
University of Technology, Sydney	4536	1955	6491	5019	1472	75%	483	25%	3064	157%
University of Western Australia	1736	1798	3534	2249	1285	71%	513	29%	451	25%
University of Western Sydney	646	496	1142	803	339	68%	157	32%	307	62%
Victoria University	1363	1218	2581	1574	1007	83%	211	17%	356	29%
	121076	82467				71%		29%		119%

University Name	Mar-05	Apr-06	Total	Non-Duplicates	Same as Previous Year	% Change	New for This Year	% Change	Lost from Previous Year	% Change
Australian Catholic University	436	434	870	568	302	70%	132	30%	134	31%
University of Adelaide	1931	1940	3871	2305	1566	81%	374	19%	365	19%
Australian National University	4718	4504	9222	5681	3541	79%	963	21%	1177	26%
University of Ballarat	250	272	522	351	171	63%	101	37%	79	29%
Bond University	258	239	497	325	172	72%	67	28%	86	36%
University of Canberra	715	610	1325	919	406	67%	204	33%	309	51%
Charles Darwin University (formally Northern Territory University)	10899	10205	21104	14382	6722	66%	3483	34%	4177	41%
Central Queensland University	1051	663	1714	1258	456	69%	207	31%	595	90%
Charles Stuart University	5711	1888	7599	6015	1584	84%	304	16%	4127	219%
Curtin University	1314	1434	2748	1712	1036	72%	398	28%	278	19%
Deakin University	1389	1225	2614	1654	960	78%	265	22%	429	35%
Edith Cowan University	1026	1525	2551	1668	883	58%	642	42%	143	9%
Flinders University	1887	2092	3979	2410	1569	75%	523	25%	318	15%
Griffith University	2038	3534	5572	4118	1454	41%	2080	59%	584	17%
James Cook University	2392	2713	5105	2809	2296	85%	417	15%	96	4%
La Trobe University	3551	3020	6571	4201	2370	78%	650	22%	1181	39%
Monash University	3212	2889	6101	4012	2089	72%	800	28%	1123	39%
Macquarie University	2522	2337	4859	3280	1579	68%	758	32%	943	40%
Murdoch University	1240	1442	2682	1795	887	62%	555	38%	353	24%
Newcastle University	1752	963	2715	2110	605	63%	358	37%	1147	119%
Queensland University of Technology	545	897	1442	1154	288	32%	609	68%	257	29%
Royal Melbourne Institute of Technology	3286	685	3971	3454	517	75%	168	25%	2769	404%
Southern Cross University	1181	668	1849	1513	336	50%	332	50%	845	126%
Swinburne University of Technology	685	31175	31860	31394	466	1%	30709	99%	219	1%
University of New England	1480	804	2284	1689	595	74%	209	26%	885	110%
University of Melbourne	2868	2934	5802	3896	1906	65%	1028	35%	962	33%
University of South Australia	1616	1446	3062	2092	970	67%	476	33%	646	45%

University of New South Wales	2692	1836	4528	3304	1224	67%	612	33%	1468	80%
University of Wollongong	1360	1238	2598	1701	897	72%	341	28%	463	37%
University of Queensland	2661	2852	5513	4220	1293	45%	1559	55%	1368	48%
University of the Sunshine Coast	186	160	346	234	112	70%	48	30%	74	46%
University of South Queensland	1549	1382	2931	1816	1115	81%	267	19%	434	31%
University of Sydney	7240	6910	14150	8147	6003	87%	907	13%	1237	18%
University of Tasmania	1359	1046	2405	1671	734	70%	312	30%	625	60%
University of Technology, Sydney	1955	1651	3606	2341	1265	77%	386	23%	690	42%
University of Western Australia	1798	1359	3157	2066	1091	80%	268	20%	707	52%
University of Western Sydney	496	324	820	556	264	81%	60	19%	232	72%
Victoria University	1218	1207	2425	1654	771	64%	436	36%	447	37%
	82467	102503				67%		33%		57%

Appendix 28: Cumulative Percentage Change in Outlinks for Australian Universities 2000 – 2006

University Name	Jan-02	Apr-06	Total	Non-Duplicates	Same as Previous Year	% Change	New for This Year	% Change	Lost from Previous Year	% Change
Australian Catholic University	708	434	1142	1003	139	32%	295	68%	569	131%
University of Adelaide	7406	1940	9346	8522	824	42%	1116	58%	6582	339%
Australian National University	4629	4504	9133	6968	2165	48%	2339	52%	2464	55%
University of Ballarat	85	272	357	336	21	8%	251	92%	64	24%
Bond University	350	239	589	535	54	23%	185	77%	296	124%
University of Canberra	489	610	1099	947	152	25%	458	75%	337	55%
Charles Darwin University (formally Northern Territory University)	2407	10205	12612	10940	1672	16%	8533	84%	735	7%
Central Queensland University	1037	663	1700	1507	193	29%	470	71%	844	127%
Charles Stuart University	7008	1888	8896	7895	1001	53%	887	47%	6007	318%
Curtin University	1815	1434	3249	2899	350	24%	1084	76%	1465	102%
Deakin University	1419	1225	2644	2179	465	38%	760	62%	954	78%
Edith Cowan University	1691	1525	3216	2117	1099	72%	426	28%	592	39%
Flinders University	1729	2092	3821	3056	765	37%	1327	63%	964	46%
Griffith University	1058	3534	4592	4264	328	9%	3206	91%	730	21%
James Cook University	2122	2713	4835	3809	1026	38%	1687	62%	1096	40%
La Trobe University	3771	3020	6791	5338	1453	48%	1567	52%	2318	77%
Monash University	3609	2889	6498	5523	975	34%	1914	66%	2634	91%
Macquarie University	2582	2337	4919	3929	990	42%	1347	58%	1592	68%
Murdoch University	2571	1442	4013	3557	456	32%	986	68%	2115	147%
Newcastle University	1677	963	2640	2302	338	35%	625	65%	1339	139%
Queensland University of Technology	1416	897	2313	2113	200	22%	697	78%	1216	136%
Royal Melbourne Institute of Technology	3701	685	4386	4172	214	31%	471	69%	3487	509%
Southern Cross University	1492	668	2160	2035	125	19%	543	81%	1367	205%
Swinburne University of Technology	649	31175	31824	31711	113	0%	31062	100%	536	2%
University of New England	1276	804	2080	1724	356	44%	448	56%	920	114%
University of Melbourne	3949	2934	6883	5957	926	32%	2008	68%	3023	103%

University of South Australia	2243	1446	3689	3277	412	28%	1034	72%	1831	127%
University of New South Wales	3002	1836	4838	4218	620	34%	1216	66%	2382	130%
University of Wollongong	1361	1238	2599	2281	318	26%	920	74%	1043	84%
University of Queensland	4565	2852	7417	6712	705	25%	2147	75%	3860	135%
University of the Sunshine Coast	70	160	230	220	10	6%	150	94%	60	38%
University of South Queensland	997	1382	2379	2041	338	24%	1044	76%	659	48%
University of Sydney	9423	6910	16333	12861	3472	50%	3438	50%	5951	86%
University of Tasmania	1640	1046	2686	2427	259	25%	787	75%	1381	132%
University of Technology, Sydney	3497	1651	5148	4547	601	36%	1050	64%	2896	175%
University of Western Australia	2305	1359	3664	3166	498	37%	861	63%	1807	133%
University of Western Sydney	885	324	1209	1124	85	26%	239	74%	800	247%
Victoria University	1210	1207	2417	2002	415	34%	792	66%	795	66%
	91844	102503	194347			31%		69%		118%

Appendix 29: Percentage Change in Outlinks for UK Universities 2000 – 2005

University Name	Jul-00	Jul-01	Total	Non-Duplicates	Same as Previous Year	% Change	New for This Year	% Change	Lost from Previous Year	% Change
University of Aberdeen	1889	4084	5973	4231	1742	43%	2342	57%	147	4%
University of Wales, Aberystwyth	2508	3229	5737	3941	1796	56%	1433	44%	712	22%
Anglia Polytechnic University	414	1978	2392	2046	346	17%	1632	83%	68	3%
Aston University	1305	1309	2614	1617	997	76%	312	24%	308	24%
University of Wales Bangor	799	1091	1890	1334	556	51%	535	49%	243	22%
University of Bath	1466	4711	6177	5267	910	19%	3801	81%	556	12%
University of Bradford	1631	2663	4294	2901	1393	52%	1270	48%	238	9%
University of Bristol	1553	8469	10022	8686	1336	16%	7133	84%	217	3%
Oxford Brookes University	372	1536	1908	1590	318	21%	1218	79%	54	4%
Brunel University	813	3485	4298	3708	590	17%	2895	83%	223	6%
University of Brighton	887	39245	40132	39424	708	2%	38537	98%	179	0%
University of Buckingham	38	58	96	61	35	60%	23	40%	3	5%
Cardiff University	2929	5852	8781	6602	2179	37%	3673	63%	750	13%
City University, London	1129	5876	7005	6133	872	15%	5004	85%	257	4%
University of Coventry	1233	1798	3031	1937	1094	61%	704	39%	139	8%
University of Derby	502	564	1066	882	184	33%	380	67%	318	56%
De Montfort University	1271	2040	3311	2393	918	45%	1122	55%	353	17%
University of Dundee	1435	2077	3512	2444	1068	51%	1009	49%	367	18%
University of Durham	3654	5472	9126	7166	1960	36%	3512	64%	1694	31%
University of Edinburgh	2643	8153	10796	9220	1576	19%	6577	81%	1067	13%
University of Essex	1710	2540	4250	3028	1222	48%	1318	52%	488	19%
University of Exeter	1896	6051	7947	6471	1476	24%	4575	76%	420	7%
Goldsmiths College, University of London	338	9488	9826	9574	252	3%	9236	97%	86	1%
University of Greenwich	798	1219	2017	1365	652	53%	567	47%	146	12%
University of Hertfordshire	2163	2768	4931	3655	1276	46%	1492	54%	887	32%
University of Huddersfield	904	1063	1967	1378	589	55%	474	45%	315	30%

Heriot-Watt University	2802	7411	10213	7608	2605	35%	4806	65%	197	3%
King’s College London	2955	4066	7021	4665	2356	58%	1710	42%	599	15%
University of Keele	1777	2318	4095	2687	1408	61%	910	39%	369	16%
Kingston University	1027	12045	13072	12116	956	8%	11089	92%	71	1%
University of Wales, Lampeter	1313	439	1752	1433	319	73%	120	27%	994	226%
University of Lincoln and Humberside	154	9814	9968	9882	86	1%	9728	99%	68	1%
University of Loughborough	1302	2569	3871	2866	1005	39%	1564	61%	297	12%
University of Leicester	2748	4782	7530	5337	2193	46%	2589	54%	555	12%
University of Leeds	3001	10844	13845	11248	2597	24%	8247	76%	404	4%
London Metropolitan University (created in August 2002 by merging the University of North London and London Guildhall University)	1239	1481	2720	1789	931	63%	550	37%	308	21%
University of Liverpool	2292	1206	3498	3360	138	11%	1068	89%	2154	179%
Liverpool John Moores University	939	1221	2160	1456	704	58%	517	42%	235	19%
Leeds Metropolitan University	665	936	1601	1103	498	53%	438	47%	167	18%
London School of Economics	552	1009	1561	1147	414	41%	595	59%	138	14%
Napier University	1220	2058	3278	2379	899	44%	1159	56%	321	16%
University of Newcastle	11	5903	5914	5904	10	0%	5893	100%	1	0%
University of Wales College Newport	158	359	517	404	113	31%	246	69%	45	13%
University of Nottingham	6427	74650	81077	76756	4321	6%	70329	94%	2106	3%
Open University	1582	3315	4897	3911	986	30%	2329	70%	596	18%
University of Paisley	797	1001	1798	1127	671	67%	330	33%	126	13%
University of Plymouth	775	11691	12466	11790	676	6%	11015	94%	99	1%
University of Portsmouth	1705	2559	4264	3283	981	38%	1578	62%	724	28%
Queen Mary, University of London	1718	4476	6194	4867	1327	30%	3149	70%	391	9%
Queen’s University Belfast	1590	3188	4778	3344	1434	45%	1754	55%	156	5%
University of Reading	4483	32492	36975	33217	3758	12%	28734	88%	725	2%
The Robert Gordon University	1013	1243	2256	1382	874	70%	369	30%	139	11%
Royal Holloway, University of London	1815	32494	34309	32674	1635	5%	30859	95%	180	1%
University of Salford	870	863	1733	1147	586	68%	277	32%	284	33%

University of Sheffield	560	7792	8352	7926	426	5%	7366	95%	134	2%
South Bank University	1625	2476	4101	2939	1162	47%	1314	53%	463	19%
School of Oriental and African Studies	211	315	526	332	194	62%	121	38%	17	5%
University of Southampton	3064	7365	10429	8215	2214	30%	5151	70%	850	12%
University of St Andrews	2457	3961	6418	4650	1768	45%	2193	55%	689	17%
University of Stirling	1485	1598	3083	1942	1141	71%	457	29%	344	22%
University of Strathclyde	2005	4050	6055	4448	1607	40%	2443	60%	398	10%
University of Sunderland	2556	7708	10264	8266	1998	26%	5710	74%	558	7%
University of Surrey	1230	2827	4057	2987	1070	38%	1757	62%	160	6%
University of Sussex	4031	36138	40169	37024	3145	9%	32993	91%	886	2%
University of Wales Swansea	2767	3096	5863	4063	1800	58%	1296	42%	967	31%
University of Teeside	352	873	1225	1043	182	21%	691	79%	170	19%
Thames Valley University	190	176	366	271	95	54%	81	46%	95	54%
University of Central England	500	660	1160	759	401	61%	259	39%	99	15%
University College London	3418	9858	13276	10569	2707	27%	7151	73%	711	7%
University of Central Lancashire	1350	1363	2713	1815	898	66%	465	34%	452	33%
University of East Anglia	1695	2282	3977	3005	972	43%	1310	57%	723	32%
University of East London	655	1006	1661	1181	480	48%	526	52%	175	17%
University of Kent at Canterbury	0	2667	2667	2660	7	0%	2660	100%	-7	0%
University of Ulster	1697	4318	6015	4575	1440	33%	2878	67%	257	6%
University of Northumbria	1648	2638	4286	3474	812	31%	1826	69%	836	32%
University of the West of England	1238	1202	2440	1810	630	52%	572	48%	608	51%
University of Wales Institute at Cardiff	425	376	801	497	304	81%	72	19%	121	32%
University of Wolverhampton	0	3383	3383	3383	0	0%	3383	100%	0	0%
University of Westminster	1204	1188	2392	1503	889	75%	299	25%	315	27%
University of York	3867	8054	11921	8936	2985	37%	5069	63%	882	11%
	123440	480622				38%		62%		19%

University Name	Jul-01	Jul-02	Total	Non-Duplicates	Same as Previous Year	% Change	New for This Year	% Change	Lost from Previous Year	% Change
University of Aberdeen	4084	4639	8723	5192	3531	76%	1108	24%	553	12%
University of Wales, Aberystwyth	3229	3637	6866	4292	2574	71%	1063	29%	655	18%
Anglia Polytechnic University	1978	3329	5307	4582	725	22%	2604	78%	1253	38%
Aston University	1309	1459	2768	1589	1179	81%	280	19%	130	9%
University of Wales Bangor	1091	1138	2229	1448	781	69%	357	31%	310	27%
University of Bath	4711	5149	9860	6433	3427	67%	1722	33%	1284	25%
University of Bradford	2663	2803	5466	3506	1960	70%	843	30%	703	25%
University of Bristol	8469	9272	17741	11777	5964	64%	3308	36%	2505	27%
Oxford Brookes University	1536	1576	3112	2063	1049	67%	527	33%	487	31%
Brunel University	3485	3552	7037	3936	3101	87%	451	13%	384	11%
University of Brighton	39245	42446	81691	46635	35056	83%	7390	17%	4189	10%
University of Buckingham	58	48	106	62	44	92%	4	8%	14	29%
Cardiff University	5852	13684	19536	16469	3067	22%	10617	78%	2785	20%
City University, London	5876	6106	11982	6610	5372	88%	734	12%	504	8%
University of Coventry	1798	1567	3365	2326	1039	66%	528	34%	759	48%
University of Derby	564	579	1143	862	281	49%	298	51%	283	49%
De Montfort University	2040	1890	3930	2398	1532	81%	358	19%	508	27%
University of Dundee	2077	2253	4330	2666	1664	74%	589	26%	413	18%
University of Durham	5472	5308	10780	7137	3643	69%	1665	31%	1829	34%
University of Edinburgh	8153	9471	17624	11455	6169	65%	3302	35%	1984	21%
University of Essex	2540	2998	5538	3493	2045	68%	953	32%	495	17%
University of Exeter	6051	6443	12494	8057	4437	69%	2006	31%	1614	25%
Goldsmiths College, University of London	9488	10703	20191	12937	7254	68%	3449	32%	2234	21%
University of Greenwich	1219	2112	3331	2314	1017	48%	1095	52%	202	10%
University of Hertfordshire	2768	3713	6481	4718	1763	47%	1950	53%	1005	27%
University of Huddersfield	1063	974	2037	1163	874	90%	100	10%	189	19%
Heriot-Watt University	7411	7178	14589	8402	6187	86%	991	14%	1224	17%

King's College London	4066	4531	8597	5270	3327	73%	1204	27%	739	16%
University of Keele	2318	1660	3978	2753	1225	74%	435	26%	1093	66%
Kingston University	12045	7974	20019	16834	3185	40%	4789	60%	8860	111%
University of Wales, Lampeter	439	410	849	518	331	81%	79	19%	108	26%
University of Lincoln and Humberside	9814	1534	11348	10166	1182	77%	352	23%	8632	563%
University of Loughborough	2569	2501	5070	3190	1880	75%	621	25%	689	28%
University of Leicester	4782	4910	9692	5931	3761	77%	1149	23%	1021	21%
University of Leeds	10844	7947	18791	12807	5984	75%	1963	25%	4860	61%
London Metropolitan University (created in August 2002 by merging the University of North London and London Guildhall University)	1481	2285	3766	2392	1374	60%	911	40%	107	5%
University of Liverpool	1206	1264	2470	1623	847	67%	417	33%	359	28%
Liverpool John Moores University	1221	1405	2626	1714	912	65%	493	35%	309	22%
Leeds Metropolitan University	936	410	1346	1004	342	83%	68	17%	594	145%
London School of Economics	1009	1292	2301	1512	789	61%	503	39%	220	17%
Napier University	2058	2024	4082	2659	1423	70%	601	30%	635	31%
University of Newcastle	5903	6240	12143	7985	4158	67%	2082	33%	1745	28%
University of Wales College Newport	359	416	775	496	279	67%	137	33%	80	19%
University of Nottingham	74650	45795	120445	100077	20368	44%	25427	56%	54282	119%
Open University	3315	3337	6652	3831	2821	85%	516	15%	494	15%
University of Paisley	1001	34	1035	1018	17	50%	17	50%	984	2894%
University of Plymouth	11691	11574	23265	20039	3226	28%	8348	72%	8465	73%
University of Portsmouth	2559	2362	4921	3057	1864	79%	498	21%	695	29%
Queen Mary, University of London	4476	3344	7820	6091	1729	52%	1615	48%	2747	82%
Queen's University Belfast	3188	3213	6401	3889	2512	78%	701	22%	676	21%
University of Reading	32492	33161	65653	50556	15097	46%	18064	54%	17395	52%
The Robert Gordon University	1243	1225	2468	1430	1038	85%	187	15%	205	17%
Royal Holloway, University of London	32494	32724	65218	37969	27249	83%	5475	17%	5245	16%
University of Salford	863	616	1479	1119	360	58%	256	42%	503	82%

University of Sheffield	7792	3549	11341	8866	2475	70%	1074	30%	5317	150%
South Bank University	2476	1899	4375	2948	1427	75%	472	25%	1049	55%
School of Oriental and African Studies	315	341	656	378	278	82%	63	18%	37	11%
University of Southampton	7365	7680	15045	9433	5612	73%	2068	27%	1753	23%
University of St Andrews	3961	23480	27441	23947	3494	15%	19986	85%	467	2%
University of Stirling	1598	1802	3400	2174	1226	68%	576	32%	372	21%
University of Strathclyde	4050	3804	7854	5014	2840	75%	964	25%	1210	32%
University of Sunderland	7708	5271	12979	8972	4007	76%	1264	24%	3701	70%
University of Surrey	2827	2653	5480	3418	2062	78%	591	22%	765	29%
University of Sussex	36138	39599	75737	59202	16535	42%	23064	58%	19603	50%
University of Wales Swansea	3096	3301	6397	4086	2311	70%	990	30%	785	24%
University of Teeside	873	625	1498	1064	434	69%	191	31%	439	70%
Thames Valley University	176	230	406	279	127	55%	103	45%	49	21%
University of Central England	660	952	1612	1155	457	48%	495	52%	203	21%
University College London	9858	6827	16685	11456	5229	77%	1598	23%	4629	68%
University of Central Lancashire	1363	1249	2612	1669	943	76%	306	24%	420	34%
University of East Anglia	2282	2439	4721	2875	1846	76%	593	24%	436	18%
University of East London	1006	758	1764	1216	548	72%	210	28%	458	60%
University of Kent at Canterbury	2667	2900	5567	3373	2194	76%	706	24%	473	16%
University of Ulster	4318	4354	8672	6592	2080	48%	2274	52%	2238	51%
University of Northumbria	2638	17502	20140	17965	2175	12%	15327	88%	463	3%
University of the West of England	1202	1750	2952	2038	914	52%	836	48%	288	16%
University of Wales Institute at Cardiff	376	293	669	495	174	59%	119	41%	202	69%
University of Wolverhampton	3383	3447	6830	3945	2885	84%	562	16%	498	14%
University of Westminster	1188	784	1972	1456	516	66%	268	34%	672	86%
University of York	8054	8817	16871	9789	7082	80%	1735	20%	972	11%
	480622	484521				66%		34%		78%

University Name	Jul-02	Jun-03	Total	Non-Duplicates	Same as Previous Year	% Change	New for This Year	% Change	Lost from Previous Year	% Change
University of Aberdeen	4639	4631	9270	5660	3610	78%	1021	22%	1029	22%
University of Wales, Aberystwyth	3637	3326	6963	4561	2402	72%	924	28%	1235	37%
Anglia Polytechnic University	3329	12764	16093	14150	1943	15%	10821	85%	1386	11%
Aston University	1459	1330	2789	1681	1108	83%	222	17%	351	26%
University of Wales Bangor	1138	1232	2370	1460	910	74%	322	26%	228	19%
University of Bath	5149	3667	8816	6446	2370	65%	1297	35%	2779	76%
University of Bradford	2803	4774	7577	5229	2348	49%	2426	51%	455	10%
University of Bristol	9272	11997	21269	14031	7238	60%	4759	40%	2034	17%
Oxford Brookes University	1576	1414	2990	1889	1101	78%	313	22%	475	34%
Brunel University	3552	3239	6791	3853	2938	91%	301	9%	614	19%
University of Brighton	42446	43013	85459	49032	36427	85%	6586	15%	6019	14%
University of Buckingham	48	53	101	59	42	79%	11	21%	6	11%
Cardiff University	13684	10160	23844	17119	6725	66%	3435	34%	6959	68%
City University, London	6106	1576	7682	6610	1072	68%	504	32%	5034	319%
University of Coventry	1567	1509	3076	1928	1148	76%	361	24%	419	28%
University of Derby	579	607	1186	690	496	82%	111	18%	83	14%
De Montfort University	1890	34	1924	1893	31	91%	3	9%	1859	5468%
University of Dundee	2253	11759	14012	12090	1922	16%	9837	84%	331	3%
University of Durham	5308	4772	10080	6297	3783	79%	989	21%	1525	32%
University of Edinburgh	9471	9306	18777	12238	6539	70%	2767	30%	2932	32%
University of Essex	2998	3233	6231	3775	2456	76%	777	24%	542	17%
University of Exeter	6443	6903	13346	7948	5398	78%	1505	22%	1045	15%
Goldsmiths College, University of London	10703	14749	25452	17471	7981	54%	6768	46%	2722	18%
University of Greenwich	2112	1659	3771	2559	1212	73%	447	27%	900	54%
University of Hertfordshire	3713	2427	6140	4070	2070	85%	357	15%	1643	68%
University of Huddersfield	974	991	1965	1122	843	85%	148	15%	131	13%
Heriot-Watt University	7178	6560	13738	7713	6025	92%	535	8%	1153	18%

King’s College London	4531	4263	8794	5489	3305	78%	958	22%	1226	29%
University of Keele	1660	2550	4210	2847	1363	53%	1187	47%	297	12%
Kingston University	7974	6078	14052	11035	3017	50%	3061	50%	4957	82%
University of Wales, Lampeter	410	272	682	408	274	101%	-2	-1%	136	50%
University of Lincoln and Humberside	1534	626	2160	1006	1154	184%	-528	-84%	380	61%
University of Loughborough	2501	2145	4646	2993	1653	77%	492	23%	848	40%
University of Leicester	4910	4420	9330	5762	3568	81%	852	19%	1342	30%
University of Leeds	7947	7830	15777	9498	6279	80%	1551	20%	1668	21%
London Metropolitan University (created in August 2002 by merging the University of North London and London Guildhall University)	2285	2226	4511	2521	1990	89%	236	11%	295	13%
University of Liverpool	1264	774	2038	1454	584	75%	190	25%	680	88%
Liverpool John Moores University	1405	1712	3117	2866	251	15%	1461	85%	1154	67%
Leeds Metropolitan University	410	376	786	468	318	85%	58	15%	92	24%
London School of Economics	1292	1500	2792	1828	964	64%	536	36%	328	22%
Napier University	2024	1901	3925	2342	1583	83%	318	17%	441	23%
University of Newcastle	6240	4436	10676	7502	3174	72%	1262	28%	3066	69%
University of Wales College Newport	416	575	991	632	359	62%	216	38%	57	10%
University of Nottingham	45795	56044	101839	72463	29376	52%	26668	48%	16419	29%
Open University	3337	3679	7016	4256	2760	75%	919	25%	577	16%
University of Paisley	34	35	69	36	33	94%	2	6%	1	3%
University of Plymouth	11574	12405	23979	18586	5393	43%	7012	57%	6181	50%
University of Portsmouth	2362	2319	4681	2801	1880	81%	439	19%	482	21%
Queen Mary, University of London	3344	2930	6274	4415	1859	63%	1071	37%	1485	51%
Queen’s University Belfast	3213	3301	6514	3714	2800	85%	501	15%	413	13%
University of Reading	33161	24074	57235	36499	20736	86%	3338	14%	12425	52%
The Robert Gordon University	1225	891	2116	1351	765	86%	126	14%	460	52%
Royal Holloway, University of London	32724	14829	47553	39915	7638	52%	7191	48%	25086	169%
University of Salford	616	1068	1684	1194	490	46%	578	54%	126	12%

University of Sheffield	3549	7122	10671	8134	2537	36%	4585	64%	1012	14%
South Bank University	1899	2209	4108	3172	936	42%	1273	58%	963	44%
School of Oriental and African Studies	341	277	618	411	207	75%	70	25%	134	48%
University of Southampton	7680	7054	14734	9071	5663	80%	1391	20%	2017	29%
University of St Andrews	23480	3469	26949	5378	21571	622%	-18102	-522%	1909	55%
University of Stirling	1802	1560	3362	2199	1163	75%	397	25%	639	41%
University of Strathclyde	3804	3445	7249	4629	2620	76%	825	24%	1184	34%
University of Sunderland	5271	3964	9235	5684	3551	90%	413	10%	1720	43%
University of Surrey	2653	2729	5382	3129	2253	83%	476	17%	400	15%
University of Sussex	39599	29379	68978	52308	16670	57%	12709	43%	22929	78%
University of Wales Swansea	3301	3403	6704	4243	2461	72%	942	28%	840	25%
University of Teeside	625	579	1204	748	456	79%	123	21%	169	29%
Thames Valley University	230	305	535	335	200	66%	105	34%	30	10%
University of Central England	952	760	1712	1210	502	66%	258	34%	450	59%
University College London	6827	8109	14936	9832	5104	63%	3005	37%	1723	21%
University of Central Lancashire	1249	1050	2299	1587	712	68%	338	32%	537	51%
University of East Anglia	2439	2302	4741	3186	1555	68%	747	32%	884	38%
University of East London	758	789	1547	982	565	72%	224	28%	193	24%
University of Kent at Canterbury	2900	35365	38265	36254	2011	6%	33354	94%	889	3%
University of Ulster	4354	8147	12501	9746	2755	34%	5392	66%	1599	20%
University of Northumbria	17502	48275	65777	54130	11647	24%	36628	76%	5855	12%
University of the West of England	1750	1484	3234	2132	1102	74%	382	26%	648	44%
University of Wales Institute at Cardiff	293	317	610	377	233	74%	84	26%	60	19%
University of Wolverhampton	3447	3178	6625	3523	3102	98%	76	2%	345	11%
University of Westminster	784	760	1544	913	631	83%	129	17%	153	20%
University of York	8817	8891	17708	9811	7897	89%	994	11%	920	10%
	484521	515866				78%		22%		105%

University Name	Jun-03	Jun-04	Total	Non-Duplicates	Same as Previous Year	% Change	New for This Year	% Change	Lost from Previous Year	% Change
University of Aberdeen	4631	4771	9402	5468	3934	82%	837	18%	697	15%
University of Wales, Aberystwyth	3326	3380	6706	4120	2586	77%	794	23%	740	22%
Anglia Polytechnic University	12764	17214	29978	20884	9094	53%	8120	47%	3670	21%
Aston University	1330	1245	2575	1477	1098	88%	147	12%	232	19%
University of Wales Bangor	1232	1116	2348	1572	776	70%	340	30%	456	41%
University of Bath	3667	3552	7219	4569	2650	75%	902	25%	1017	29%
University of Bradford	4774	4923	9697	6477	3220	65%	1703	35%	1554	32%
University of Bristol	11997	20867	32864	24780	8084	39%	12783	61%	3913	19%
Oxford Brookes University	1414	1449	2863	1821	1042	72%	407	28%	372	26%
Brunel University	3239	2844	6083	3620	2463	87%	381	13%	776	27%
University of Brighton	43013	28217	71230	64193	7037	25%	21180	75%	35976	127%
University of Buckingham	53	60	113	67	46	77%	14	23%	7	12%
Cardiff University	10160	17150	27310	19563	7747	45%	9403	55%	2413	14%
City University, London	1576	1337	2913	1834	1079	81%	258	19%	497	37%
University of Coventry	1509	1095	2604	1825	779	71%	316	29%	730	67%
University of Derby	607	405	1012	700	312	77%	93	23%	295	73%
De Montfort University	34	20	54	49	5	25%	15	75%	29	145%
University of Dundee	11759	11674	23433	13186	10247	88%	1427	12%	1512	13%
University of Durham	4772	4513	9285	5705	3580	79%	933	21%	1192	26%
University of Edinburgh	9306	9335	18641	11521	7120	76%	2215	24%	2186	23%
University of Essex	3233	3086	6319	4056	2263	73%	823	27%	970	31%
University of Exeter	6903	6398	13301	8695	4606	72%	1792	28%	2297	36%
Goldsmiths College, University of London	14749	9334	24083	17227	6856	73%	2478	27%	7893	85%
University of Greenwich	1659	1667	3326	2030	1296	78%	371	22%	363	22%
University of Hertfordshire	2427	2206	4633	2633	2000	91%	206	9%	427	19%
University of Huddersfield	991	684	1675	1083	592	87%	92	13%	399	58%
Heriot-Watt University	6560	5787	12347	7156	5191	90%	596	10%	1369	24%

King's College London	4263	10549	14812	11534	3278	31%	7271	69%	985	9%
University of Keele	2550	2069	4619	3107	1512	73%	557	27%	1038	50%
Kingston University	6078	7631	13709	11067	2642	35%	4989	65%	3436	45%
University of Wales, Lampeter	272	343	615	400	215	63%	128	37%	57	17%
University of Lincoln and Humberside	626	540	1166	635	531	98%	9	2%	95	18%
University of Loughborough	2145	2088	4233	2690	1543	74%	545	26%	602	29%
University of Leicester	4420	4833	9253	5473	3780	78%	1053	22%	640	13%
University of Leeds	7830	7756	15586	9882	5704	74%	2052	26%	2126	27%
London Metropolitan University (created in August 2002 by merging the University of North London and London Guildhall University)	2226	31472	33698	32661	1037	3%	30435	97%	1189	4%
University of Liverpool	774	1188	1962	1371	591	50%	597	50%	183	15%
Liverpool John Moores University	1712	1548	3260	2170	1090	70%	458	30%	622	40%
Leeds Metropolitan University	376	1329	1705	1533	172	13%	1157	87%	204	15%
London School of Economics	1500	1634	3134	1953	1181	72%	453	28%	319	20%
Napier University	1901	1635	3536	2179	1357	83%	278	17%	544	33%
University of Newcastle	4436	7096	11532	8247	3285	46%	3811	54%	1151	16%
University of Wales College Newport	575	463	1038	729	309	67%	154	33%	266	57%
University of Nottingham	56044	77622	133666	103546	30120	39%	47502	61%	25924	33%
Open University	3679	2655	6334	4492	1842	69%	813	31%	1837	69%
University of Paisley	35	1051	1086	1077	9	1%	1042	99%	26	2%
University of Plymouth	12405	4322	16727	14761	1966	45%	2356	55%	10439	242%
University of Portsmouth	2319	2291	4610	3046	1564	68%	727	32%	755	33%
Queen Mary, University of London	2930	3460	6390	4784	1606	46%	1854	54%	1324	38%
Queen's University Belfast	3301	2360	5661	3761	1900	81%	460	19%	1401	59%
University of Reading	24074	9492	33566	27056	6510	69%	2982	31%	17564	185%
The Robert Gordon University	891	837	1728	994	734	88%	103	12%	157	19%
Royal Holloway, University of London	14829	6244	21073	16065	5008	80%	1236	20%	9821	157%
University of Salford	1068	964	2032	1484	548	57%	416	43%	520	54%
University of Sheffield	7122	8540	15662	10242	5420	63%	3120	37%	1702	20%

South Bank University	2209	1638	3847	2781	1066	65%	572	35%	1143	70%
School of Oriental and African Studies	277	227	504	307	197	87%	30	13%	80	35%
University of Southampton	7054	6140	13194	8299	4895	80%	1245	20%	2159	35%
University of St Andrews	3469	119125	122594	120359	2235	2%	116890	98%	1234	1%
University of Stirling	1560	1684	3244	2147	1097	65%	587	35%	463	27%
University of Strathclyde	3445	2888	6333	4060	2273	79%	615	21%	1172	41%
University of Sunderland	3964	3590	7554	4837	2717	76%	873	24%	1247	35%
University of Surrey	2729	2420	5149	3268	1881	78%	539	22%	848	35%
University of Sussex	29379	51016	80395	61060	19335	38%	31681	62%	10044	20%
University of Wales Swansea	3403	3821	7224	4693	2531	66%	1290	34%	872	23%
University of Teeside	579	467	1046	669	377	81%	90	19%	202	43%
Thames Valley University	305	340	645	433	212	62%	128	38%	93	27%
University of Central England	760	776	1536	919	617	80%	159	20%	143	18%
University College London	8109	8411	16520	10057	6463	77%	1948	23%	1646	20%
University of Central Lancashire	1050	982	2032	1351	681	69%	301	31%	369	38%
University of East Anglia	2302	2259	4561	2902	1659	73%	600	27%	643	28%
University of East London	789	807	1596	1001	595	74%	212	26%	194	24%
University of Kent at Canterbury	35365	13694	49059	14148	34911	255%	-21217	-155%	454	3%
University of Ulster	8147	9983	18130	13008	5122	51%	4861	49%	3025	30%
University of Northumbria	48275	21962	70237	53695	16542	75%	5420	25%	31733	144%
University of the West of England	1484	1468	2952	2099	853	58%	615	42%	631	43%
University of Wales Institute at Cardiff	317	414	731	473	258	62%	156	38%	59	14%
University of Wolverhampton	3178	3170	6348	3773	2575	81%	595	19%	603	19%
University of Westminster	760	770	1530	1004	526	68%	244	32%	234	30%
University of York	8891	5110	14001	9993	4008	78%	1102	22%	4883	96%
	515866	629503				68%		32%		41%

University Name	Jun-04	Jul-05	Total	Non-Duplicates	Same as Previous Year	% Change	New for This Year	% Change	Lost from Previous Year	% Change
University of Aberdeen	4771	2894	7665	5321	2344	81%	550	19%	2427	84%
University of Wales, Aberystwyth	3380	2976	6356	3898	2458	83%	518	17%	922	31%
Anglia Polytechnic University	17214	1996	19210	17810	1400	70%	596	30%	15814	792%
Aston University	1245	932	2177	1397	780	84%	152	16%	465	50%
University of Wales Bangor	1116	1112	2228	1320	908	82%	204	18%	208	19%
University of Bath	3552	3968	7520	4551	2969	75%	999	25%	583	15%
University of Bradford	4923	4155	9078	6442	2636	63%	1519	37%	2287	55%
University of Bristol	20867	28107	48974	38425	10549	38%	17558	62%	10318	37%
Oxford Brookes University	1449	1356	2805	1761	1044	77%	312	23%	405	30%
Brunel University	2844	696	3540	3170	370	53%	326	47%	2474	355%
University of Brighton	28217	28723	56940	36252	20688	72%	8035	28%	7529	26%
University of Buckingham	60	69	129	71	58	84%	11	16%	2	3%
Cardiff University	17150	11269	28419	22843	5576	49%	5693	51%	11574	103%
City University, London	1337	1156	2493	1664	829	72%	327	28%	508	44%
University of Coventry	1095	896	1991	1152	839	94%	57	6%	256	29%
University of Derby	405	360	765	556	209	58%	151	42%	196	54%
De Montfort University	20	707	727	714	13	2%	694	98%	7	1%
University of Dundee	11674	11686	23360	12103	11257	96%	429	4%	417	4%
University of Durham	4513	4102	8615	5448	3167	77%	935	23%	1346	33%
University of Edinburgh	9335	11810	21145	13480	7665	65%	4145	35%	1670	14%
University of Essex	3086	3294	6380	3769	2611	79%	683	21%	475	14%
University of Exeter	6398	6591	12989	8384	4605	70%	1986	30%	1793	27%
Goldsmiths College, University of London	9334	9968	19302	12917	6385	64%	3583	36%	2949	30%
University of Greenwich	1667	1433	3100	2026	1074	75%	359	25%	593	41%
University of Hertfordshire	2206	1987	4193	2463	1730	87%	257	13%	476	24%
University of Huddersfield	684	741	1425	925	500	67%	241	33%	184	25%
Heriot-Watt University	5787	6259	12046	6618	5428	87%	831	13%	359	6%

King’s College London	10549	10862	21411	11574	9837	91%	1025	9%	712	7%
University of Keele	2069	2283	4352	3009	1343	59%	940	41%	726	32%
Kingston University	7631	9968	17599	12877	4722	47%	5246	53%	2909	29%
University of Wales, Lampeter	343	367	710	411	299	81%	68	19%	44	12%
University of Lincoln and Humberside	540	194	734	610	124	64%	70	36%	416	214%
University of Loughborough	2088	2303	4391	2859	1532	67%	771	33%	556	24%
University of Leicester	4833	5437	10270	6314	3956	73%	1481	27%	877	16%
University of Leeds	7756	8161	15917	9676	6241	76%	1920	24%	1515	19%
London Metropolitan University (created in August 2002 by merging the University of North London and London Guildhall University)	31472	19000	50472	34710	15762	83%	3238	17%	15710	83%
University of Liverpool	1188	869	2057	1311	746	86%	123	14%	442	51%
Liverpool John Moores University	1548	1680	3228	1994	1234	73%	446	27%	314	19%
Leeds Metropolitan University	1329	1477	2806	2268	538	36%	939	64%	791	54%
London School of Economics	1634	1758	3392	2295	1097	62%	661	38%	537	31%
Napier University	1635	1376	3011	1862	1149	84%	227	16%	486	35%
University of Newcastle	7096	4193	11289	8023	3266	78%	927	22%	3830	91%
University of Wales College Newport	463	606	1069	718	351	58%	255	42%	112	18%
University of Nottingham	77622	288287	365909	332974	32935	11%	255352	89%	44687	16%
Open University	2655	2666	5321	3178	2143	80%	523	20%	512	19%
University of Paisley	1051	1172	2223	1249	974	83%	198	17%	77	7%
University of Plymouth	4322	4223	8545	5222	3323	79%	900	21%	999	24%
University of Portsmouth	2291	2377	4668	2785	1883	79%	494	21%	408	17%
Queen Mary, University of London	3460	5224	8684	6513	2171	42%	3053	58%	1289	25%
Queen’s University Belfast	2360	2790	5150	3302	1848	66%	942	34%	512	18%
University of Reading	9492	12959	22451	15434	7017	54%	5942	46%	2475	19%
The Robert Gordon University	837	441	1278	931	347	79%	94	21%	490	111%
Royal Holloway, University of London	6244	7510	13754	9367	4387	58%	3123	42%	1857	25%
University of Salford	964	1079	2043	1362	681	63%	398	37%	283	26%

University of Sheffield	8540	9605	18145	12537	5608	58%	3997	42%	2932	31%
South Bank University	1638	1355	2993	1806	1187	88%	168	12%	451	33%
School of Oriental and African Studies	227	55	282	253	29	53%	26	47%	198	360%
University of Southampton	6140	6288	12428	7631	4797	76%	1491	24%	1343	21%
University of St Andrews	119125	48466	167591	140150	27441	57%	21025	43%	91684	189%
University of Stirling	1684	1718	3402	2086	1316	77%	402	23%	368	21%
University of Strathclyde	2888	3108	5996	3818	2178	70%	930	30%	710	23%
University of Sunderland	3590	3058	6648	4014	2634	86%	424	14%	956	31%
University of Surrey	2420	2199	4619	2815	1804	82%	395	18%	616	28%
University of Sussex	51016	17036	68052	51784	16268	95%	768	5%	34748	204%
University of Wales Swansea	3821	3927	7748	4883	2865	73%	1062	27%	956	24%
University of Teeside	467	161	628	528	100	62%	61	38%	367	228%
Thames Valley University	340	305	645	433	212	70%	93	30%	128	42%
University of Central England	776	692	1468	930	538	78%	154	22%	238	34%
University College London	8411	8895	17306	10416	6890	77%	2005	23%	1521	17%
University of Central Lancashire	982	1102	2084	1282	802	73%	300	27%	180	16%
University of East Anglia	2259	2093	4352	2756	1596	76%	497	24%	663	32%
University of East London	807	875	1682	1049	633	72%	242	28%	174	20%
University of Kent at Canterbury	13694	10287	23981	15125	8856	86%	1431	14%	4838	47%
University of Ulster	9983	33501	43484	37567	5917	18%	27584	82%	4066	12%
University of Northumbria	21962	654	22616	22150	466	71%	188	29%	21496	3287%
University of the West of England	1468	1824	3292	2104	1188	65%	636	35%	280	15%
University of Wales Institute at Cardiff	414	414	828	494	334	81%	80	19%	80	19%
University of Wolverhampton	3170	3244	6414	3747	2667	82%	577	18%	503	16%
University of Westminster	770	701	1471	942	529	75%	172	25%	241	34%
University of York	5110	5191	10301	6265	4036	78%	1155	22%	1074	21%
	629503	71525 9				70%		30%		97%

Appendix 30: Cumulative Percentage Change in Outlinks for UK Universities 2000 – 2005

University Name	Jul-01	Jul-05	Total	Non-Duplicates	Same as Previous Year	% Change	New for This Year	% Change	Lost from Previous Year	% Change
University of Aberdeen	4084	2894	6978	5833	1145	40%	1749	60%	2939	102%
University of Wales, Aberystwyth	3229	2976	6205	5005	1200	40%	1776	60%	2029	68%
Anglia Polytechnic University	1978	1996	3974	3619	355	18%	1641	82%	1623	81%
Aston University	1309	932	2241	1740	501	54%	431	46%	808	87%
University of Wales Bangor	1091	1112	2203	1844	359	32%	753	68%	732	66%
University of Bath	4711	3968	8679	7328	1351	34%	2617	66%	3360	85%
University of Bradford	2663	4155	6818	6268	550	13%	3605	87%	2113	51%
University of Bristol	8469	28107	36576	33590	2986	11%	25121	89%	5483	20%
Oxford Brookes University	1536	1356	2892	2521	371	27%	985	73%	1165	86%
Brunel University	3485	696	4181	4013	168	24%	528	76%	3317	477%
University of Brighton	39245	28723	67968	64651	3317	12%	25406	88%	35928	125%
University of Buckingham	58	69	127	92	35	51%	34	49%	23	33%
Cardiff University	5852	11269	17121	14499	2622	23%	8647	77%	3230	29%
City University, London	5876	1156	7032	6580	452	39%	704	61%	5424	469%
University of Coventry	1798	896	2694	2186	508	57%	388	43%	1290	144%
University of Derby	564	360	924	838	86	24%	274	76%	478	133%
De Montfort University	2040	707	2747	2408	339	48%	368	52%	1701	241%
University of Dundee	2077	11686	13763	12717	1046	9%	10640	91%	1031	9%
University of Durham	5472	4102	9574	7731	1843	45%	2259	55%	3629	88%
University of Edinburgh	8153	11810	19963	16140	3823	32%	7987	68%	4330	37%
University of Essex	2540	3294	5834	4574	1260	38%	2034	62%	1280	39%
University of Exeter	6051	6591	12642	10184	2458	37%	4133	63%	3593	55%
Goldsmiths College, University of London	9488	9968	19456	17602	1854	19%	8114	81%	7634	77%
University of Greenwich	1219	1433	2652	2157	495	35%	938	65%	724	51%
University of Hertfordshire	2768	1987	4755	3598	1157	58%	830	42%	1611	81%
University of Huddersfield	1063	741	1804	1404	400	54%	341	46%	663	89%

Heriot-Watt University	7411	6259	13670	9442	4228	68%	2031	32%	3183	51%
King’s College London	4066	10862	14928	12943	1985	18%	8877	82%	2081	19%
University of Keele	2318	2283	4601	4019	582	25%	1701	75%	1736	76%
Kingston University	12045	9968	22013	21449	564	6%	9404	94%	11481	115%
University of Wales, Lampeter	439	367	806	648	158	43%	209	57%	281	77%
University of Lincoln and Humberside	2514	3394	5908	5390	518	15%	2876	85%	1996	59%
University of Loughborough	2569	2303	4872	4064	808	35%	1495	65%	1761	76%
University of Leicester	4782	5437	10219	7923	2296	42%	3141	58%	2486	46%
University of Leeds	10844	8161	19005	15954	3051	37%	5110	63%	7793	95%
London Metropolitan University (created in August 2002 by merging the University of North London and London Guildhall University)	1481	19000	20481	19883	598	3%	18402	97%	883	5%
University of Liverpool	1206	869	2075	1676	399	46%	470	54%	807	93%
Liverpool John Moores University	1221	1680	2901	2325	576	34%	1104	66%	645	38%
Leeds Metropolitan University	936	1477	2413	2338	75	5%	1402	95%	861	58%
London School of Economics	1009	1758	2767	2400	367	21%	1391	79%	642	37%
Napier University	2058	1376	3434	2781	653	47%	723	53%	1405	102%
University of Newcastle	5903	4193	10096	8351	1745	42%	2448	58%	4158	99%
University of Wales College Newport	359	606	965	885	80	13%	526	87%	279	46%
University of Nottingham	74650	288287	362937	355807	7130	2%	281157	98%	67520	23%
Open University	3315	2666	5981	5076	905	34%	1761	66%	2410	90%
University of Paisley	1001	1172	2173	2045	128	11%	1044	89%	873	74%
University of Portsmouth	2559	2377	4936	3941	995	42%	1382	58%	1564	66%
Queen’s University Belfast	3188	2790	5978	5027	951	34%	1839	66%	2237	80%
University of Reading	32492	12959	45451	43457	1994	15%	10965	85%	30498	235%
Royal Holloway, University of London	32494	7510	40004	36185	3819	51%	3691	49%	28675	382%
University of Salford	863	1079	1942	1733	209	19%	870	81%	654	61%
University of Sheffield	7792	9605	17397	15160	2237	23%	7368	77%	5555	58%
South Bank University	2476	1355	3831	3127	704	52%	651	48%	1772	131%

School of Oriental and African Studies	315	55	370	366	4	7%	51	93%	311	565%
University of Southampton	7365	6288	13653	11157	2496	40%	3792	60%	4869	77%
University of St Andrews	3961	48466	52427	5727	46700	96%	1766	4%	-42739	-88%
University of Stirling	1598	1718	3316	2879	437	25%	1281	75%	1161	68%
University of Strathclyde	4050	3108	7158	5923	1235	40%	1873	60%	2815	91%
University of Sunderland	7708	3058	10766	9012	1754	57%	1304	43%	5954	195%
University of Surrey	2827	2199	5026	4059	967	44%	1232	56%	1860	85%
University of Sussex	36138	17036	53174	49934	3240	19%	13796	81%	32898	193%
University of Wales Swansea	3096	3927	7023	5518	1505	38%	2422	62%	1591	41%
University of Teeside	873	161	1034	994	40	25%	121	75%	833	517%
Thames Valley University	176	305	481	428	53	17%	252	83%	123	40%
University of Central England	660	692	1352	1018	334	48%	358	52%	326	47%
University College London	9858	8895	18753	15059	3694	42%	5201	58%	6164	69%
University of Central Lancashire	1363	1102	2465	2061	404	37%	698	63%	959	87%
University of East Anglia	2282	2093	4375	3599	776	37%	1317	63%	1506	72%
University of East London	1006	875	1881	1639	242	28%	633	72%	764	87%
University of Kent at Canterbury	2667	10287	12954	11585	1369	13%	8918	87%	1298	13%
University of Ulster	4318	33501	37819	36323	1496	4%	32005	96%	2822	8%
University of Northumbria	2638	654	3292	2923	369	56%	285	44%	2269	347%
University of the West of England	1202	1824	3026	2702	324	18%	1500	82%	878	48%
University of Wales Institute at Cardiff	376	414	790	682	108	26%	306	74%	268	65%
University of Wolverhampton	3383	3244	6627	4673	1954	60%	1290	40%	1429	44%
University of Westminster	1188	701	1889	1609	280	40%	421	60%	908	130%
University of York	8054	5191	13245	10972	2273	44%	2918	56%	5781	111%
	455912	708571		1023993		33%		67%		104%

Appendix 31: Top 10 UK Inter-University Domain Links 2001 - 2004

2001			2002			2003			2004		
From	To	No of Links	From	To	No of Links	From	To	No of Links	From	To	No of Links
Cambridge	Oxford	724	Cambridge	Oxford	615	Cambridge	Oxford	701	Cambridge	Oxford	642
Oxford	Cambridge	559	Oxford	Cambridge	457	Oxford	Cambridge	442	Oxford	Cambridge	445
Cambridge	Imperial College London	283	Edinburgh	Cambridge	275	Glasgow	Edinburgh	282	University College London	Cambridge	266
University College London	Cambridge	279	Cambridge	Imperial College London	241	Cambridge	Edinburgh	280	Cambridge	Edinburgh	260
Glasgow	Edinburgh	274	Glasgow	Edinburgh	237	Edinburgh	Cambridge	278	Edinburgh	Cambridge	250
Cambridge	Edinburgh	272	University College London	Cambridge	224	Cambridge	Imperial College London	274	Glasgow	Edinburgh	244
South bank University	Oxford	267	Cambridge	Edinburgh	222	University College London	Cambridge	274	Cambridge	Imperial College London	231
University College London	Oxford	244	Edinburgh	Oxford	221	Edinburgh	Oxford	234	Edinburgh	Glasgow	224
Edinburgh	Cambridge	236	Edinburgh	Glasgow	203	Cambridge	University College London	228	Edinburgh	Oxford	217
Glasgow	Oxford	224	University College London	Oxford	203	Edinburgh	Glasgow	227	Cambridge	University College London	216